Temperament Traits in Cattle: Measurement and Preliminary Genetic Analysis

Nia Ball

Doctor of Philosophy

The University of Edinburgh

2003

Acknowledgements

I would like to acknowledge my supervisors, Marie Haskell, John Williams, John Deag and Mike Appleby, and thank them for their help on the farm and reading through drafts of chapters. I particularly thank Marie for her help and advice. The studentship was funded by the Roslin Institute.

Many colleagues willingly helped me move cattle around for behaviour experiments, even though most clearly thought I was mad. I would like to thank Wilson Lee, Laura Nicoll and the rest of the staff at Blythbank Farm, and Kenny Rutherford, Anita Rennie, Helena Pickup, Nancy Coerse, Polly Taylor and Martin Yeates.

Statistical advice throughout the project was gratefully received from Anthea Springbett. Making sense of the genotyping data involved help from many people, some who kindly helped out at short notice. In particular I would like to thank Deb Burton, Pam Wiener, Andy Law, Chris Haley, Pau Navarro and Richard Papworth. Chris Haley and George Seaton allowed me to use an unreleased software program, and Pierre Le Neindre from INRA kindly lent me a program of his. Norrie Russell and Elliot Armstrong have helped out with photographs and posters.

Many colleagues in the Institute have made my time at Roslin very enjoyable, especially the Witches of East Wing and their cronies (whose teatime stories have been an education) and the Hut 11 group. Friends in Edinburgh have shared many rants, laughs and bottles of gin over the last few years, and their support has been very much appreciated.

Finally, special thanks are due to my parents, Neris and Hywel Ball, for their endless support and encouragement, and their many words of wisdom.

Abstract

Individual differences in fearful and aggressive behaviours are seen in farm animal species. Behavioural responses of cattle to handling and management systems on farms are presumed to reflect underlying temperament traits such as fearfulness. As temperament measurements have a strong genetic component, selection for desirable behavioural responses could increase the ability of animals to cope with stressors encountered on farms. The aims of this project were i) to take measures of behaviour in cattle that could be demonstrated to be reliable indicators of temperament traits, and ii) to look for areas of the genome harbouring genes involved in these traits.

Behavioural tests were carried out on a large number of animals from a Charolais x Holstein cross-bred herd. Four different tests were examined for their potential to measure different temperament traits: a Flight-from-Feeder Test (FF), a Social Separation Test (SS), a Novel Object Test (NO) and a Handling Test (HA). The results of the tests were assessed using two criteria: inter-animal variability and intra-animal repeatability. The FF, SS and HA Tests showed high values in these two criteria. Experiments were then carried out to validate these three tests, by examining whether relationships were seen between responses from different test situations that were thought to measure the same traits. Full validation of the three tests was not achieved.

Experiments were also carried out to investigate how stable the behaviour measurements remained with increasing age. Results of SS and HA Tests carried out at four months of age were compared with those from the same tests repeated on the animals at 12 months of age. A relationship was found between SS measures from the two ages. Results from behavioural tests carried out on heifers at 10 months of age were compared with behaviour scored in the dairy parlour at 30 months. The measurements at the younger age were not predictive of adult behaviour in the dairy.

In order to localise areas of the genome involved in the behavioural traits, preliminary Quantitative Trait Loci (QTL) analysis was carried out on four chromosomes. Associations between behavioural phenotypes from the FF and SS Tests and the inheritance of DNA markers were examined for a large number of animals. The possible effects of environmental factors on the test results were examined prior to the analysis. Four putative QTL locations were identified.

This study has demonstrated that reliable measures of temperament can be made in cattle, and has added to the information available on the genetic control of behaviour in cattle.

Contents

Chapter One: Introduction

1.1. Introduction

The focus of this thesis is the measurement of fearfulness traits in cattle, and the investigation of the genetic control of the traits. This chapter introduces this subject. It describes the differences in temperament traits that are seen between individual animals, which are of especial interest in farm animal species. Fearfulness in particular is an important trait, because of its potential impact on the welfare of animals. After describing stimuli that cause fearfulness in cattle, the use of behavioural tests to measure fearfulness of three common stimuli is then described.

The latter sections of the chapter describe the extent of genetic influences on temperament traits in animals, and detail the evidence for significant genetic effects on the differences in fearfulness traits seen between individual cattle. The recent discovery of abundant DNA markers has made possible the search for the location of genes that affect quantitative traits such as behaviour. Here, the method of using microsatellite markers to locate quantitative trait loci (QTLs) is described, along with the requirements necessary for such a search. Studies that have identified QTLs for temperament traits in cattle and in mice are discussed.

Finally, the aims of this project were to measure fearfulness traits in cattle and identify QTLs for those traits. These aims are described in detail in the final section of the chapter.

1.2. Temperament Traits in Animals

Differences between individuals in how they react to stimuli have been measured in a wide range of species, from chimpanzees to guppies (Gosling, 2001). The behavioural responses of an animal to an external stimulus are modified by a central state of subjective experience. The term 'temperament' is commonly used to refer to this central emotional state, which has motivational properties (Goldsmith *et al.*, 1987). Differences between individual animals in their behavioural responses are presumed to reflect differences in the perception and/or cognitive evaluation of the stimulus (Ramos & Mormède, 1998). For example, differences in behaviour between individuals may be seen when a new toy is introduced to a litter of kittens. One kitten may approach the object immediately to touch and push it, while its sibling may approach cautiously, sniffing the object, and preparing to retreat quickly if it does anything unexpected. It is assumed that the differences in the behavioural styles of these two individuals reflect differences in their phenomenological experience. The first kitten is motivated to explore or to play with the object, whereas the second appears to be fearful of the object. Such differences result from characteristics of the animals' neural system, which are as yet undefined (Lawrence *et al.*, 1991). Behavioural traits (broad behavioural consistencies) shown in a range of situations are thought to be reflective of temperament traits, and, if used sensibly, may provide us with an indirect measure of the subjective experiences of animals.

Temperament is typically thought of as having an inborn, genetically influenced, biological basis (Goldsmith, 1989), although it is also affected by sex and age, and modified by the environment. Environmental effects include maternal effects and, in the case of domestic animals, rearing conditions and early handling of the animals. The environment can continue to modify temperament throughout life, though to a lesser extent.

This way of thinking about how observable behaviour patterns can reflect underlying temperament traits in animals is borrowed from a long history of research by psychologists into human personality. The distinction between 'temperament' and 'personality' is fuzzy. Strelau (1983) differentiated between the two terms in several ways, including the distinction that temperament can be seen in both animals and humans, whereas personality describes the phenomena that are specific to humans, such as 'ego' and 'self', or certain cognitive constructs (Zuckerman, 1991). Researchers continue to investigate broad behavioural consistencies seen in the conduct of people that are thought to represent basic categories of individual differences in functioning (Pervin, 1996). By measuring people on a large number of personality characteristics, factor analysis methods are used to determine whether the many differences seen between people boil down to a few groupings. Since the 1990s there have been many such studies carried out (Pervin, 1996). A consensus seems to be slowly emerging that there are five basic factors or dimensions of personality (the 'Five Factor' model; McCrae & Costa, 1985). These factors have been named 'Neuroticism', 'Extraversion', 'Conscientiousness', 'Agreeableness' and 'Openness to Experience'.

A large number of traits in each of the five broad factors are used to characterise someone's personality, such as outgoing, friendly, reserved, hostile or competitive. Traits are measured in people using three main methods: tests of behaviour, behaviour ratings (both self-ratings and ratings by others), and questionnaires about their behaviours. The use of self-report and questionnaires are not methods that can be borrowed for assessing temperament traits in animals, but ratings of behaviour observed in free situations and direct recording of discrete behaviours displayed in test situations are often used. These methods are described later in Section 1.5. They have been used to measure a range of temperament traits in many species of animals, for example: responses to novelty in rhesus monkeys (Stevenson-Hinde *et al.*, 1980), fearfulness of humans in mink (Hansen, 1996), emotionality in mice (Flint *et al.*, 1995), activity in cats (Feaver *et al.*, 1986), affability in dogs (Wilsson & Sundgren, 1997), timidity in goats (Lyons *et al.*, 1988), responses to handling in pigs (Lawrence *et al.*, 1991), aggression in cattle (Sánchez *et al.*, 1996) and sociality in chicks (Marin *et al.*, 2001).

Temperament traits in farm animals are of great interest for two main reasons, concerns about animal welfare, and ease of management of animals. Greater attention has been paid to improving welfare as the potential impact of poor welfare on levels of production has become clear. Exposure to stressful stimuli is unavoidable in any modern farm system. Although challenges from the environment are natural, if too many occur within a short period of time, or a chronic stressor is present, welfare is likely to be compromised (Wemelsfelder & Birke, 1997). Hence, temperament research in farm animals is usually focussed on the measurement of traits leading to negative emotions, such as fear. Fearful animals which try to avoid humans can be difficult to handle, leading to increased handling time and risk of injury of both stockpersons and the animals during routine inspections, especially when dealing with large species such as cattle. There is much still to learn about temperament traits in cattle, and fearfulness in this species is the focus of this thesis. The trait of fearfulness and how it is measured in animals is discussed next.

1.3. Fearfulness in Animals

The temperament trait of fearfulness corresponds to emotional states of fear (perception of actual danger) and anxiety (perception of potential danger). It is defined as the general propensity of an individual to perceive and react in the same manner to a wide variety of potentially threatening events (Boissy, 1998). Events that can cause fear in domestic animals include novelty, contact with humans, sudden movement, disruption of the social environment (such as separation from familiar companions, or exposure to strangers), specific stimuli that relate to persistent dangers in the ecology of the species, and learned threatening events. The term emotionality is often used as a synonym of fearfulness.

Fear is such a powerful emotion that it effectively competes with and progressively inhibits behaviour patterns generated by all other motivational systems (Jones, 1996). Because fear takes precedence over other behaviour systems, we can infer how frightened an animal is by monitoring its responses in test situations intuitively regarded as more or less frightening (Jones, 1996). A wide variety of behavioural tests have been developed over the last century to measure fearfulness, anxiety and emotionality in rodents, and a huge number of studies has been carried out, using procedures that expose animals to either physically or psychologically aversive stimuli. Many of these tests are now also widely used in farm animal research, and some are described in Section 1.5.

The interest in individual differences in temperament has led researchers to ask whether individuals are consistent in their responses to *different* challenges (Boissy, 1995). Indeed, it is assumed in the definition given above that an animal characterised as fearful is predisposed to show more intense and prolonged behavioural and physiological fear reactions when exposed to *any* non-specific disturbance and potential danger (Jones, 1996). These two authors regard fearfulness as a unidimensional trait. The alternative view, that fearfulness cannot be regarded as one simple trait in this way, was first aired by Archer (1973), when he reviewed the measurement of fearfulness in novel environment tests in mice and rats. He found that emotionality measures taken from a variety of tests showed little or no relationship to one another, and suggested that such results 'clearly failed to support the use of emotionality as a consistent constituent trait, with unitary drive properties'.

Boissy (1995) believed the results reviewed by Archer could be explained by procedural variations between studies and differences between individual animals, and did not believe that they represented evidence of a lack of unidimensionality of the fearfulness trait. Strong intra-individual correlations found between scores in different fearfulness tests in domestic hens led Jones (1996) to suggest that the tests do measure the same intervening variable, presumably underlying non-specific fearfulness, in this species (Jones, 1987; Jones, 1988; Jones & Waddington, 1992). Similarly, correlations between behavioural patterns shown by many animals (including cattle and sheep) in various aversive situations led Boissy (1998) to conclude that 'there is a strong tendency for individual characteristics to manifest themselves across a variety of aversive situations'.

Ramos and Mormède (1998) recently reviewed a large number of studies of anxiety in rodents, and found inconsistencies in behavioural correlations between different studies and different strains. They concluded that 'either one (or all) of the tests considered is unable to produce a pure and reliable measure of anxiety, or else that the different tests assess different forms of anxiety', a conclusion similar to that reached by Archer (1973) over 20 years previously.

Hence whether fearfulness can be regarded as a single trait remains to be fully answered. Ramos and Mormède (1998) argue convincingly that the trait is multidimensional, certainly in rodents, but whether this is also the case in other species such as cattle remains to be demonstrated. The position adopted at the start of this project was that fearfulness is likely to be composed of a set of traits in cattle. Taking this stance, and measuring reactions to different stimuli as different traits during the first experiments carried out, allowed this issue to be investigated. Therefore reactions to the different stimuli that cause fearfulness in cattle are discussed as separate traits in this review.

1.4. Fearfulness in Cattle

During their lifetime, cattle are likely to encounter a range of stimuli that are likely to elicit fear and distress. Three of the most common are contact with humans, novel stimuli or environments and disruption of the social environment (especially separation from familiar conspecifics). These three causes of fear will be concentrated on in this thesis. Others include certain husbandry procedures, transportation, and pre-slaughter handling.

Many farming systems, especially dairy systems, impose a high level of contact between cattle and people, involving tactile, visual and olfactory stimuli. Human contact is encountered on a daily basis for husbandry procedures such as feeding, bedding and moving animals, and close handling takes place when animals are milked or undergo examinations or veterinary procedures. The frequency of such interactions and the use of routine behaviours by stockpeople can result in cattle becoming highly fearful of humans. Research has shown that these high fear levels, through stress, appear to limit animal productivity and welfare (Hemsworth & Coleman, 1998). Studies carried out on dairy farms by Rushen *et al.* (1999) and Breuer *et al.* (2000) suggest that negative handling may depress the milk yield of cows. A further study showed that improving the attitudinal and behavioural profiles of stockpersons towards their animals can lead to a reduction in the level of fear and increase in the productivity of dairy cows (Hemsworth *et al.*, 2002). As a consequence of such research, the model shown in Figure 1.1 has been proposed by Hemsworth & Coleman (1998) to describe the influence of human-animal interactions on the productivity and welfare of animals.

Exposure to novelty is a powerful method of inducing and examining fear responses in animals (Boissy, 1998). Cattle exhibit strong signs of fear of unfamiliar stimuli (neophobia). This can make handing difficult when animals are exposed to unfamiliar equipment or a new housing arrangement. Social isolation is another potent source of fear. Individual animals differ in the extent to which they need social companionship, or in their 'sociality'. It is not clear to what extent sociality can be considered as a fearfulness trait. Cattle are a gregarious species (excluding mature bulls), and imposing isolation upon individuals can cause severe behavioural responses (Kilgour, 1975; Boissy & Le Neindre, 1997). When separated from the herd unexpectedly, extreme reactions are often seen, such as attempting to jump over gates to rejoin the herd. Hence, here I consider sociality to be a fearfulness trait. Some regard the behaviour of animals in social isolation as a combination of their sociality (motivation to end the social isolation) and their particular reaction pattern when in a stressful situation (Erhard & Schouten, 2001). During routine husbandry practices cattle are often separated temporarily for procedures such as artificial insemination and foot-paring, and a high level of sociality exhibited by animals makes handling more difficult in these cases. The presence of peers can reduce stress responses of cattle to other fear-eliciting stimuli, such as distress behaviour of heifers in response to a novel object (Boissy & Le Neindre, 1990) and responses to human handling (Grignard *et al.*, 2000).

How temperament traits are measured in animals, and specifically how responses to different frightening stimuli are measured in cattle, are discussed in the next section.

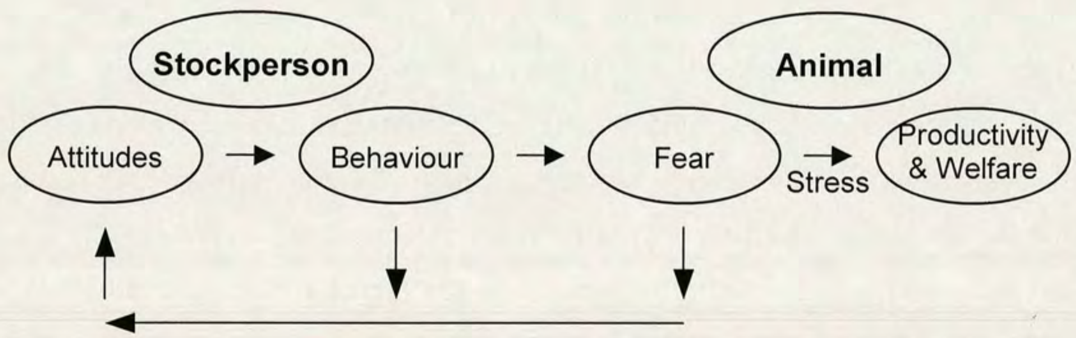


Figure 1.1: Model of human-animal interactions (Hemsworth & Coleman, 1998).

1.5. Testing Fearfulness in Cattle

1.5.1. Behavioural Tests in Animals

1.5.2. Tests used in Cattle

1.5.1. Behavioural Tests in Animals

Ratings of behaviour observed in free situations and direct recording of discrete behaviours displayed in test situations are used to measure temperament in a wide range of animals. A method of rating observers' evaluations has been applied to rhesus monkeys (Stevenson-Hinde *et al.*, 1980) and domestic cats (Feaver *et al.*, 1986). This method provides ratings of overall patterns of behaviour observed in free situations, rather than discrete events (Martin & Bateson, 1993). Each animal is rated on a number of categories, which are adjectives such as active, curious, aggressive or playful. The method requires the human observer to be thoroughly familiar with the individual animals (Martin & Bateson, 1993).

Behavioural tests have also been used extensively, and this is normally the method used in farm animal species, as a large number of animals can be assessed without the need for the observers to be familiar with the individual animals. Behavioural tests are highly controlled and allow discrete behavioural responses to be measured and recorded as objectively as possible. They facilitate the comparison of individuals in a very standardised way, which allows the effects of temperament to become clearly evident, as different responses are evoked in different individuals by the same environmental context (Lyons, 1989). The results are usually easy to analyse, and can sometimes be compared between studies.

However, care must be taken, as the behaviour measured during the test is a result not only of the temperament of the individual, but of an interaction between its temperament and also its psychobiological state at the time (which depends on factors such as nutritional state and social status), and the precise test conditions (Mormède *et al.*, 2002). To ensure that responses measured during the test are being determined largely by the animal's temperament, and the effects of its emotional state at that precise time are negligible, consistency in behaviour measures across repeated tests needs to be assessed. To determine which elements of the animal's temperament are being brought out by the precise conditions of the test, the measures used need to be carefully validated, by examining them for cross-situational consistency. These two criteria are discussed.

1.5.1.1. Repeatability

The concept of temperament traits implies that the traits included are relatively consistent across time, and so the behavioural responses examined need to be demonstrated to be more than momentary emotional responses to particular situations (Goldsmith *et al.*, 1987). The term 'state' can be used for the behaviour an animal performs or the mood it is in at a particular moment in time in a specific situation, but only if the individual is repeatedly found to be in similar states can we assume we are measuring an underlying trait (Erhard & Schouten, 2001). Test-retest repeatability, a measure of the degree to which measures can be replicated, should be demonstrated for each test procedure used. Many studies of individual differences in behaviour carried out on farm animals have performed tests only once on each animal, with no attempt to check whether the test procedure gives consistent results when repeated on the same animals. This lack of evidence of repeatability makes interpretation of the reactions seen difficult.

1.5.1.2. Interpretation and Validity

Validity concerns the extent to which behavioural measurements actually measure those traits the investigator wishes to measure (Manteca & Deag, 1993). Measures of fear (or other temperament traits) are inferred from behaviour seen during tests. Often interpretations used have been over-simplistic and the measures taken lack evidence of validity (Rushen, 2000). The main method of validating behavioural results is to look at whether animals express the same traits in other similar situations (cross-situational consistency; Pervin, 1996). Correlations found between behaviours in different tests show that the behavioural responses are useful measures, and not purely a result of very specific responses to the immediate test environment only (Mendl & Harcourt, 1988). However, it is arguable that relationships between just test situations, which may be arbitrary, provide strong validation, and, whenever possible, attempts should be made to validate the behaviour seen in test situations with behaviour shown in 'real-life' situations.

Temperament is thought to be composed of several dimensions. Consistency or lack of it in one of these realms is potentially independent of consistency or lack of it in the other (Goldsmith *et al.*, 1987). Therefore, correlations are only expected between behaviours from tests that are thought to measure the same trait. As discussed in section 1.3, it is now believed that fearfulness may be composed of a set of traits. Hence fearful reactions shown in one situation may not correlate with fearful reactions shown in another if the fear-causing stimuli are different.

Some of the behavioural tests used to assess animals' reactions to these three stimuli are discussed in the following section. Possible genetic influences on the responses seen in such tests are discussed in section 1.6.2.

1.5.2. Tests used in Cattle

1.5.2.1. Human Contact

Burrow (1997) reviewed many tests that have been used to measure the behavioural responses of cattle to humans, and to handling by humans. The tests could be classed into restrained tests (where the animal's movements are physically restricted in a crush or weighing scale), and non-restrained tests (during which the animal is free to move within a relatively large test area in the presence or absence of an observer). Burrow (1997) concluded that behaviours related to an animal's fear response of humans are best identified by non-restrained tests, as it is not always possible to relate behaviours in a restrained situation to behaviours in a non-restrained situation. I will therefore concentrate on two non-restrained tests that have been widely used and found to give useful measures of fearfulness of humans - flight tests and the Docility Test.

Flight tests measure the behaviour of an animal when actively approached by a human being. Usually a measure of 'flight distance' is taken. The observer, who is unfamiliar to the animals, slowly walks up to an animal until it starts to walk away, and notes the distance between them at this point. Flight test responses by cattle have been measured in a range of different experimental conditions, ranging from testing the animals at pasture when in close proximity to herdmates (Murphey *et al.*, 1980;

1981) to testing in pens (Kabuga & Appiah, 1992; Dennison, 1986) or yards when the animals are isolated (Fordyce *et al.*, 1982; Fisher *et al.*, 2000). The posture of the observer and their direction of approach relative to the animal's body also often differ between studies. Because the animals' responses can be affected by the observer's posture and movement, and the presence of conspecifics (Manteca & Deag, 1993), comparisons between the results of different experiments must be made with care.

The Docility Test described by Boivin and his colleagues (1992a) has been used extensively to evaluate differences between animals in their behavioural reactions to human handling. In the test each animal is separated from its herdmates in turn and an attempt is made to restrain it in a corner of a pen. A 'Docility Score', a continuous variable integrating the different behavioural responses seen, is calculated (Le Neindre *et al.*, 1995). The test has been used to measure ease of handling in the investigation of the effects of rearing conditions and handling of calves at an early age on their subsequent behaviour (Boivin *et al.*, 1992a,b; Boivin *et al.*, 1994), and the effects of breed and sire on reactions to handling procedures (Boivin *et al.*, 1994); Le Neindre *et al.*, 1995; Grignard *et al.*, 2001).

1.5.2.2. Novelty

Response to novelty is commonly measured in two main ways, using a Novel Object Test, or an Open Field Test. The Open Field Test also incorporates other stimuli including social separation, and is discussed below. During a Novel Object Test, an animal is typically placed in a pen for a set amount of time with an object that it has not experienced before, and its reactions towards the object are recorded. Novel objects used have included colourful striped inflatable balls (Murphey *et al.*, 1981; Hemsworth *et al.*, 1996), an iron truncated pyramid with white and green stripes (Boissy & Bouissou, 1995), a white plastic bucket and a white chair (Hemsworth *et al.*, 1996), traffic cones (Boissy *et al.*, 1998; Plusquellec & Bouissou, 2001) and a red rotating light placed upon a bucket under which a buzzer rang continuously (Plusquellec & Bouissou, 2001). Measures such as 'latency to approach' and 'time spent away from' the object are used to estimate fearfulness, as it is assumed that a fearful animal will take longer to approach the object, and spend less time in contact with it.

1.5.2.3. Sociality

Individual differences in sociality are often tested using the Open Field (OF) Test. During an OF Test, the animal is placed in a novel enclosure and its behavioural responses are monitored. The OF test has been used in a wide range of studies, for example, to compare activity motivation in calves from housing systems with differing degrees of physical restriction (Dellmeier *et al.*, 1990), and to assess whether calves that perform stereotypic behaviours perceive stressful situations differently from those that don't (Redbo, 1998).

The big disadvantage of using the OF Test as a method of assessing sociality is the number of factors that contribute to the animals' responses in addition to the social isolation. These factors include the novelty of the environment, dimensions and features of the room itself, the effects of handling by humans prior to the test, the opportunity for locomotion, and the opportunity for exploration (Munksgaard & Jensen, 1996). Each of these factors varies in extent between experimental set-ups. For example, testing arenas may vary in how different they are from the home environment, and range from outdoor arenas (e.g. Dellmeier *et al.*, 1990) to enclosed rooms (e.g. Boissy & Bouissou, 1994). The impact of isolation may be different in group-reared animals and individually-reared or isolated animals (Munksgaard & Jensen, 1996).

The range of factors that are involved, and the fact that they all differ in degree between experiments, makes the behavioural responses seen during the tests difficult to interpret. In an attempt to unpick the factors affecting behavioural responses, researchers have experimentally varied the suspected influences one at a time, noted changes in responses, and made interpretations accordingly. Lack of agreement between different interpretations is frequent. For example, some authors have interpreted locomotory/ambulatory measures as reflective of the animals' internal motivation for locomotion (Dellmeier *et al.*, 1990; Jensen, 2001). They found that more locomotory behaviour occurred during the tests if the animals' opportunity to perform such behaviours had previously been restricted. However, de Passillé *et al.* (1995) found no effect of previous exercise, and Veissier & Le Neindre (1992) suggested that these behaviours are reflective of escape attempts. De Passillé *et al.* (1995) also considered running and walking separately, as each seemed to reflect different motivations; an internal drive for locomotion and response to novelty, respectively.

Despite the variety of results and interpretations, individual differences in OF Test responses are found (Kilgour, 1975; Boissy & Bouissou, 1988), and the responses have been found to be consistent when the test is repeated on the same animals (Kilgour, 1975; Hopster & Blokhuis, 1993). Therefore the test seems to measure a fairly consistent mix of traits in each case. However, because of the variety of influences, the OF Test is unlikely to be useful for the investigation of single temperament traits. Tests that use a similar set-up but which cut down on some of the factors involved may be very useful, for example, social separation tests which are carried out in the animals' familiar home environment. Hopster and Blokhuis (1993) carried out such a test. They imposed isolation on one dairy cow per day, by letting all animals out to pasture after milking as usual apart from the test animal, which was held in a familiar cubicle house for an hour.

Other methods of testing sociality have been used but are much less common. Social motivation has been measured using a set-up where the animal has to clear a frightening obstacle (Plusquelle & Bouissou, 2001) or a barrier (Matthews *et al.*, 1997) to rejoin its penmates. However the reactions to these tests are potentially confounded by individual differences in perception of the object that needs to be overcome to reach their penmates. Boissy and Le Neindre (1997) looked at the responses of heifers to isolation without the confounding effects of handling or a novel environment by carrying out lengthy habituation procedures. They saw strong reactions to separation, but as the animals were restrained in a crush during the test, physical restraint is likely to have played a large part in the response. Fisher *et al.* (2000) carried out a 'sociability test', in which the time taken for each animal to move from the top of a yard to join a group of conspecifics 30 m away was measured.

The selected examples given in this section have demonstrated the large number of different temperament tests that have been carried out, looking at the reactions of cattle to a variety of fearful stimuli. Many studies have shown a large influence of genetics on fearfulness and other temperament traits. These studies are described in the following section.

1.6. Genetic Influences on Fearfulness

1.6.1. Genetics of Temperament Traits in Animals

Heredity has a large effect on the temperament of wild and domestic animals. Indeed, it is the extent of the genetic influence on behaviour that has allowed the successful domestication of animals. Most domestic animals have been artificially selected to have a lower reactivity to man and sudden changes in their environment than their wild ancestors (Newman, 1994).

Heritability (in the narrow sense of the term) is the proportion of phenotypic differences among individuals that can be attributed to additive genetic variance (Plomin *et al.*, 1997). If the heritability of a trait is greater than zero, it is possible to increase or decrease the extent of the trait phenotype by selection, and the higher the heritability, the greater the response to selection will be.

Evidence of the effect of genetics on the variability of temperament traits between animals from the same species comes from two main sources. The first is in the differences seen between animals from different breeds, and the second is selection studies. Dogs provide a good example of a species that shows extreme differences between breeds. As well as great variability in size and appearance between difference breeds, there are also big behavioural effects, as a result of artificial selection in their breeding history. For example, terriers were bred to creep into burrows to drive out small animals, and are aggressive scrappers, whereas spaniels were bred to point, and are non-aggressive and people-orientated (Plomin *et al.*, 1997).

In typical selection studies, high and low lines of a laboratory animal are selected in addition to the maintenance of an unselected control line (Plomin *et al.*, 1997). Successful selection can only occur if heredity is important. In one selection study of behaviour, mice were selected for activity in the open-field test, where lower activity scores are presumed to index fearfulness (DeFries *et al.*, 1978). The most active mice were selected and mated with other high-active mice, and the least active mice were also mated with each other. This selection process was repeated for 30 generations. Over the generations, selection was successful: the high lines became increasingly more active and the low lines less active. After 30 generations a thirtyfold average difference in activity was achieved (Plomin *et al.*, 1997).

It is difficult to study the genetics of behaviour in cattle compared to many other animals, because of the long generation interval and small number of offspring produced. Therefore, most of the evidence of the important role that genetics plays in behavioural traits in cattle comes from differences seen between breeds. Selection for behavioural traits has been practiced since humans started to domesticate cattle (Buchenauer, 1999). Different breeds have undergone behaviour-selection for different traits. It is evident that selection for behavioural characteristics played a very important part in the early stages of the development of milk and beef breeds (Albright & Arave, 1997). Dairy breeds have undergone considerable genetic selection to facilitate milking (Albright & Arave, 1997), and dairy cows are generally more docile than beef cows (although dairy bulls are much more aggressive than beef bulls). Some breeds have been selected by breeders for their tendency to fight, such as the Hérens, an alpine dairy breed whose cows fight when unfamiliar animals meet. Animals are pitted against each other in traditional shows (Plusquellec & Bouissou, 2001).

Buchenauer (1999) reviewed the genetics of social behaviour, temperament, sexual behaviour, maternal behaviour and feeding behaviour in cattle, and concluded that most of the traits showed heritability values large enough to allow the traits to respond to selection. The genetics of temperament traits are of particular interest, especially those of fearfulness. As discussed in Section 1.4, high levels of fear in cattle cause stress and can limit productivity. If heritability values for fearfulness traits are high, a knowledge of the genetics of these traits would allow such measures to be incorporated into selective breeding programmes. This has the potential to genetically improve behavioural responses of animals, increasing adaptability of the animals to farm environments, and improving ease of handling. Hence the genetic factors are of economic and ethical significance for farm animals (Boissy *et al.*, 2002).

Boissy and his colleagues (2002) reviewed the genetics of fearfulness in domestic herbivores. They concluded that there was evidence of differences in fear responses between breeds when cattle were exposed to humans, open-field tests, surprise tests and novel object tests. Burrow (1997) summarised heritabilities of different types of tests of the responses of beef cattle to different tests of handling by humans, and concluded that temperament measures in beef cattle were at least moderately heritable, indicating that they should response to selection.

Below, I have focussed on the evidence for genetic effects on the fearfulness traits in cattle discussed in Section 1.5.2.

1.6.2. Genetics of Fearfulness in Cattle

1.6.2.1. Contact with Humans

Many different tests have been carried out to measure the reactions of cattle to humans. The two which appear to be the most useful are flight tests and the Docility Test, and these were described in Section 1.5.2.1. Evidence of genetic effects on the responses of animals to contact with humans is provided by breed differences found in response to flight tests, and by the calculation of a heritability estimate of the Docility Score from the Docility Test.

Murphey *et al.* (1980) investigated the flight distance of cows of 12 breeds from 18 herds, which were reared in similar open pasture systems. The breeds included both European humpless *(Bos taurus)* and Zebu humped *(Bos indicus)* breeds, some of which were kept as milk animals and some for beef production. They found that some breeds of cattle were more approachable than others, and clear evidence that dairy cows were more approachable than beef cows. Although the herds were raised on different farms, a comparison of three breeds reared as dairy cows at one farm and as beef at another showed that breed differences accounted for much more of the variation in flight distance than rearing treatment, and that dairy breeds were more approachable than beef breeds even when beef breeds were treated as milk cows. Hence it appears from this study that genetic factors have a large effect on flight distance.

Breed differences in flight distance have also been found between Angus and Limousin bulls, which are both beef breeds (Vanderwert *et al.*, 1985). As the animals from the two breeds were acquired from different sources in this experiment, it is unclear to what extent the differences were due to genetics rather than environmental effects. Contrary to these two studies, Fordyce *et al.* (1982) found *no* evidence of differences between breed groups when they measured flight distance in bulls and heifers from three beef *B. taurus* and *B. indicus* breeds and their crossbreds. There were also no differences in flight distance shown by cows and heifer calves from the N'dama (a beef breed) and West African Shorthorn (a breed kept for ceremonial purposes), or between N'dama and Holstein (a dairy breed; Kabuga & Appiah, 1992). The results of these two studies suggest that genetic influences on flight distance may be small. However, the breed differences were also confounded by rearing differences in these studies.

Although the Docility Test has been carried out on animals from different breeds, there is no evidence for breed differences in the docility scores obtained. No differences in scores between three groups of artificially-reared heifers of different dairy breeds (Tarine, Montbeliarde and Friesian) were found (Boivin *et al.*, 1992a). Grignard and colleagues (2000) examined calves from two different beef breeds (Salers and Limousin) but made no between-breed comparison. However, Le Neindre and his colleagues (1995) carried out the test on a large number of Limousin (beef) heifers, in order to estimate the genetic variability of docility within the herd. They estimated a heritability value for the Docility Score of 0.22, high enough to select efficiently for the trait. As the heifers came from many different farms, differences in handling before weaning may have contributed to variability in docility, and therefore true heritability values may be higher than these estimates.

1.6.2.2. Novelty

Few studies have looked at breed differences in response to novel objects. No clear difference was found between cows from two dairy breeds in their response to a cone or a rotating light (Plusquelle & Bouissou, 2001). Murphey *et al.* (1981) investigated differences in responses to a striped ball between 21 herds of 12 *B. taurus*, *B. indicus* and *B. taurus* x *B. indicus* breeds in Brazil, including cows from both beef and dairy breeds. The herds were ranked according to their investigatory behaviour of the ball. This suggests that differences were seen between herds, but no formal statistical tests were carried out to investigate differences between them. It is unclear whether the differences reflected previous husbandry experiences of the different herds or true breed differences.

1.6.2.3. Sociality

Breed differences in open-field and sociality behaviour have been found in a few of the studies mentioned in section 1.5.2.3. Differences in open-field behaviour were seen between Salers (beef) and Friesian (dairy) cows that had been subject to the same rearing conditions (Le Neindre, 1989). However, in most of the experiments, the differences between breed groups were confounded by differences in rearing practices. Boissy and Le Neindre (1997) found differences in sociality between Friesian heifers reared in a dairy system and Aubrac (dairy) heifers reared under suckler conditions. Plusquelle and Bouissou (2001) found only tendencies towards differences between cows from two alpine dairy breeds. Details of the rearing conditions of these animals were unclear.

1.6.2.4. Summary

The experiments described above, although not numerous, have shown that the influence of genetics can be important in responses to at least two of these three stimuli that cause fear in cattle. Clear differences in responses to humans have been demonstrated to exist between beef and dairy breeds, and a heritability estimate has been calculated for a measure of this trait. Clear differences have also been seen between dairy and beef breeds in open-field behaviour, which is thought to incorporate, among other factors, a sociality response. Other studies have also demonstrated differing responses between different dairy breeds, and between different beef breeds, for both these traits. Unfortunately all these studies used animals from different rearing backgrounds, making it impossible to conclude that differences seen were definitely due to genetic effects. However, it seems likely that they were, at least in part.

Recently, molecular tools have become available which makes it possible to search for areas of the genome that influence traits such as these. Two studies have identified chromosomal regions where genes that influence temperament traits in cattle are located, using a method called 'quantitative trait loci (QTL) analysis'

(Schmutz *et al.*, 2001; Fisher *et al.*, 2001). The principles of this type of analysis, and the tools required for such a study, are described in the following section.

1.7. QTL Analysis

1.7.1. Introduction
1.7.2. Requirements for QTL Detection
1.7.3. QTLs for Behavioural Traits

1.7.1. Introduction

Over recent years, there has been increasing interest in the genetics of 'quantitative' production traits in farm animals. Some traits of importance to animal breeding are qualitative traits which determined by just one single gene, for example, muscular hypertrophy, or 'double muscling' in cattle (Nicholas, 2000). Single trait genes are investigated relatively easily using an understanding of simple Mendelian inheritance. However, most traits of economic importance are quantitative (Simm, 1998), and involve the combined action of many genes.

A quantitative trait is one which shows a continuous distribution in phenotype, such as milk production or body weight, and these are usually under the influence of several genes. Each gene in such a multiple-gene system is called a quantitative trait locus (QTL; Plomin *et al.*, 1997). The recent discovery of abundant DNA markers has made it possible to find the approximate locations of QTLs responsible for quantitative traits. Markers are identifiable segments of DNA whose positions in the genome are known, and which have distinct alleles that can easily be followed through generations of a pedigree. If a QTL lies in close physical proximity to a DNA marker on a chromosome, the frequency of recombination events separating them during meiosis will be very low, and the two will tend to be inherited together. This linkage between QTLs and DNA markers allows the marker to act as a proxy for the QTL, and allows the entire genome to be searched for the locations of genes affecting traits of interest.

To identify QTLs affecting a quantitative trait, a large number of individual animals are needed. Each animal is scored for their alleles at a few hundred DNA markers distributed throughout the genome (their genotype), and for their phenotype for the trait. If a QTL affecting the trait lies within the same chromosomal region as one of the markers, then the phenotype and marker will be frequently inherited together, and the association between them can be detected. Markers linked to QTLs can be used in selection programmes, even when the coding sequence of the gene itself is still unknown. Genotypes for each linked marker are used as additional clues to the true breeding value of an animal for a trait in marker-assisted selection (MAS) programmes (Nicholas, 2000).

Hence the requirements for detecting QTLs are firstly, a map of marker loci that covers the whole genome, and secondly, a population of animals that show variation for the quantitative phenotype of interest. These two requirements are discussed in 1.7.2, which focuses mainly on the use of microsatellite DNA markers, and a population structure created by crossing two parental breeds that are divergent in phenotype. These are the most common tools used to identify QTLs in farm animals, and are the methods used in this thesis. Also detailed are statistical methods used for the detection of the location and mode of action of QTLs. Recent findings of QTLs for behavioural traits in mice and in cattle are summarised in 1.7.3.

1.7.2. Requirements for QTL Detection

1.7.2.1. Microsatellite Markers and Maps

Any identifiable segment of DNA in the genome, which shows variation between animals, can be used as a marker (Simm, 1998). The location of each marker within the genome is known, allowing each marker to act as a DNA 'signpost', marking out specific regions of chromosomes. The discovery of microsatellite sequences has provided an abundant source of markers, which has allowed the construction of DNA marker maps in a wide range of mammalian species in the last ten years (Georges, 1998).

Microsatellites consist of tandem repeats of nucleotide sequences of between one and six bases long, for example, CACACACACACACACA. They have multiple alleles, which vary between individual animals in their number of repeats (Weber & May, 1989). For example, one animal may have a $(CA)_8$ allele at a certain marker locus, while another may have $(CA)_{10}$ or $(CA)_{14}$. This variation in allele size seen between individuals is called polymorphism. It is the high level of polymorphism of microsatellite loci and the large number of them distributed evenly throughout the genome that makes them so important for use as markers (Queller *et al.*, 1993).

Laboratory techniques are used to identify which alleles each animal has at each marker locus. If sequences adjacent to the repeated sections are known, they can be used to prime the synthesis of new strands of DNA. The polymerase chain reaction (PCR) is used to amplify a microsatellite locus from each individual. When separated by size on an electrophoresis gel, the number of repeats (or the allele size) can be determined for each animal, allowing each to be assigned to one of many different genotypes.

Collaborative studies involving many laboratories have taken place to identify the locations of large numbers of microsatellite loci and construct marker maps. In cattle, marker maps are constructed using material from shared 'reference families' which have been selected to be optimal for map construction (Georges, 1998). Several cattle families have been used for the construction of genetic linkage maps in cattle, using crosses between divergent breeds (Bishop *et al.*, 1994; Barendse *et al.*, 1993; 1994; 1997). Large full-sib families are created using multiple ovulation embryo transfer, and half-sib families are produced by mating a sire to a large number of dams using artificial insemination.

Thirty-six laboratories contributed to the Cattle Genotypic Database, which is a fast access database for information on the cattle genome (http://www.cgd.csiro.au). The database has combined raw genotypes from all the families to create the International Bovine Reference Panel (IBRP) consensus maps. Combined efforts have resulted in more than 1600 microsatellite marker loci located in all regions of the bovine genome (Barendse & Fries, 1999). A map composed of a few hundred of these markers can be used to roughly locate the QTLs for quantitative traits.

Map distances between markers are expressed by the number of recombination events that occur between them. A one per cent recombination rate is equivalent to a one centiMorgan (cM) genetic distance. There is no universal relationship between map distances calculated this way and the actual physical distance between loci. A centiMorgan can correspond to a span of between 10,000 and 1,000,000 nucleotide base pairs in mammals (Lynch & Walsh, 1998). In most experiments, a map distance of ~20 cM is the limit of resolution, so what is detected as a QTL is a segment of chromosome of this length, which may contain several loci affecting the trait.

1.7.2.2. Population Structure

The high polymorphism of microsatellite loci allows the inheritance of alleles from parent to offspring to be traced through a pedigree. When experimental herds are set up, a population comprised of three generations (F_0 grandparents, F_1 parents and F_2 offspring) is used, from which genotypic data is required from the F_2 individuals and their parents and grandparents, and phenotypic data is required from the F_2 animals only.

Figure 1.2 shows an example of inheritance, through three generations, of a marker linked to a QTL affecting a trait of interest (size) in two breeds of pigs. In this example the two breeds differ in size and also for markers distributed through the genome. The breeds are inbred lines, so each grandparent is homozygous ('1 1' or '2 2') for this example marker. Haley and Andersson (1997) explain the principles of QTL mapping as follows:

> "The F_1 animals inherit one chromosome from each breed and hence one copy of each allele of the marker (i.e. they are '1 2' heterozygotes). In the F_2 generation (produced by crossing two F_1 animals) the marker segregates giving '1 1' and '2 2' homozygotes and '1 2' heterozygotes. Because of genetic linkage, animals inheriting the marker allele from one breed also tend to inherit the surrounding region of chromosome from the same breed.
>
> There is also segregation in the F_2 for the genes that control the size differences between the two breeds, including the QTL we are seeking. In this example we find, in the F_2 animals, an association between the marker alleles they inherit and their size: those with the '1 1' marker genotype tend to be larger than those with the '2 2' genotype, while '1 2' animals are intermediate in size. This association shows that one or more QTLs which contribute size differences between the breeds must be linked (and therefore close) to the marker."

This situation, a cross between two inbred lines, is the simplest, as each line is homozygous for both the marker and the QTL. In cattle, inbred lines like this are not available. Instead it is assumed that long established breeds which have very different selection histories, for example beef and dairy breeds, each have unique QTLs of moderate or large effect (Haley *et al.*, 1994), and these are crossed to create experimental resource herds. Resource herds provide powerful experimental tools, as the breeding, management and trait measurements of the animals can be strictly controlled.

Individuals from the two divergent breeds are mated and the F_1 individuals are used to generate a large F_2 or backcross population (Georges, 1998). The inheritance of chromosomal segments from parents to offspring is again followed using DNA markers spanning the genome, and for this the F_1 parents must be heterozygous (Barendse & Fries, 1999). It is slightly more complicated than a cross between inbred lines, because only a proportion of the F_1 animals will be heterozygous for a given marker locus. This proportion depends on how polymorphic the marker is in the particular population, and is called the 'informativeness' of the marker. Additionally, marker alleles can have different associations with different QTLs in different sire families. Therefore the analysis carried out needs to track the inheritance of marker alleles from the purebred animals through the first and second generation crosses within each family (Haley & Andersson, 1997). One of the most commonly used statistical methods that has been developed for this is 'interval mapping'.

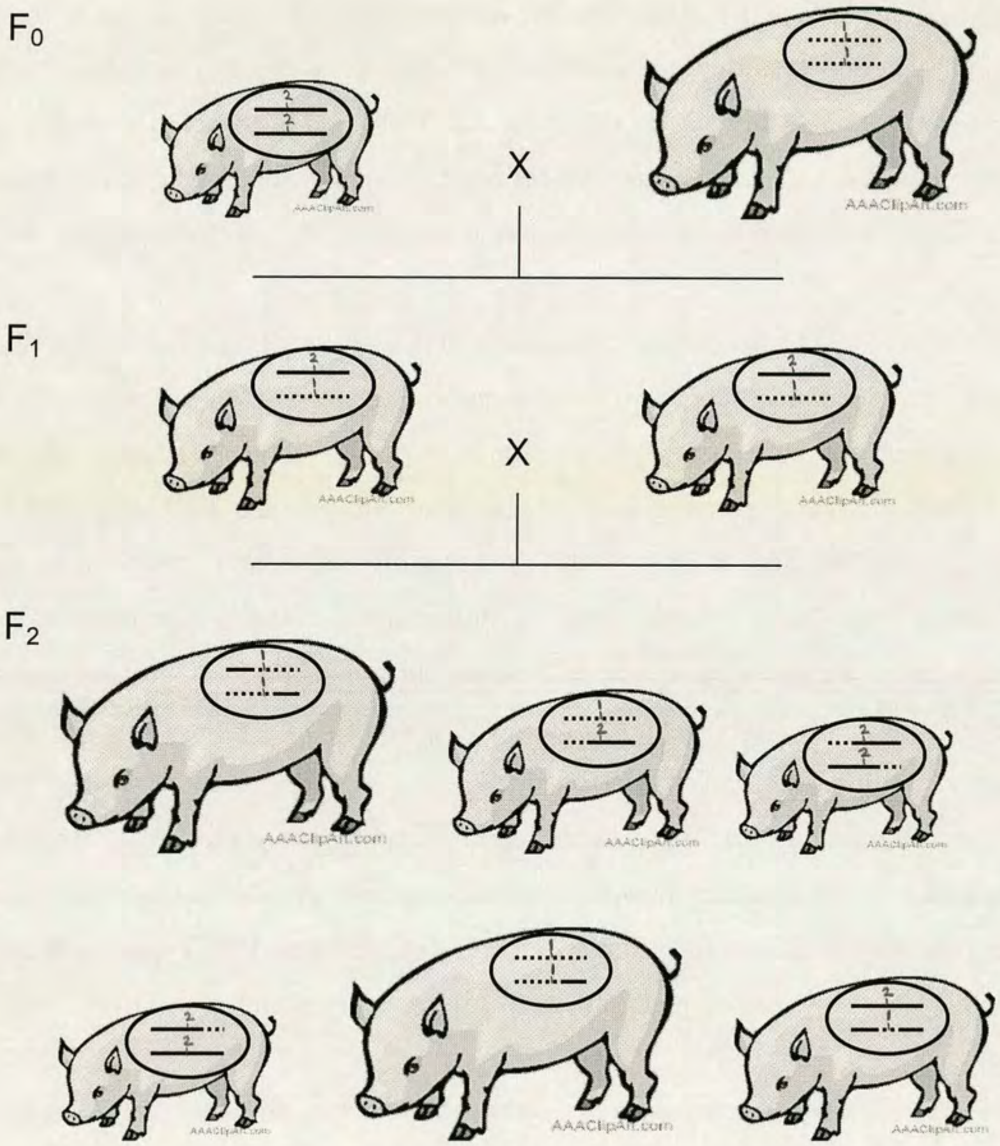


Figure 1.2: The principle of QTL detection using a genetic marker. Adapted from Haley and Andersson (1997).

1.7.2.3. QTL Probability Calculation

Interval mapping analysis is a method of estimating the position and effect of a QTL, and it uses data from all markers simultaneously. Compared to methods which consider only one single marker at a time, interval mapping methods have been shown to provide additional power and more accurate estimates of QTL effect and position (Lander & Botstein, 1989). Haley and his colleagues (1994) developed a multiple regression method that allows a least squares analysis to be applied to the interval mapping of QTLs in a cross between outbred lines or breeds. The method involves two stages. Firstly, the markers are used to follow inheritance of alleles from parents to offspring, as in Figure 1.2, and probabilities of the different genotypes occurring at each of many locations throughout the genome are obtained. Then, a statistical model is used to regress the phenotypes of the F_2 animals with the genotype probabilities at each location. The evidence for the presence of a QTL at each location is presented as a test statistic, and the position at which a QTL would explain most of the variation between marker alleles is identified.

The calculation of a significance threshold is then necessary, to indicate how strong the evidence is for the presence of the QTL. Permutation tests can be carried out to set significance thresholds at the single-point or chromosome-wide level (Churchill & Doerge, 1994). Permutation tests involve repeat 'shuffling' of the quantitative trait values and the generation of a random sample of the test statistic under the null hypothesis that there are no QTL segregating (Churchill & Doerge, 1994). Significance thresholds are obtained from this distribution (Haley & Andersson, 1997). The threshold can be set at different levels, and is a trade-off between too stringent and too lax a standard (Lander & Kruglyak, 1995). A single-position threshold estimates separately for each analysis point, of which there are many on each chromosome. If these are used, significant results will be obtained by chance (type 1 error). A chromosome-wide threshold provides an overall critical value that is valid for all points on a chromosome, and is higher than the single-position threshold. It is used to detect the presence of a QTL somewhere on the chromosome, while controlling the overall type 1 error rate (Churchill & Doerge, 1994). Even using a chromosome-wide level of significance at 5 % means that 1.5 QTLs per whole-genome scan may be expected to occur by chance, and hence when a whole-genome scan is being carried out, a higher, genome-wide significance threshold should be used. Therefore the price for controlling type 1 error over the entire genome is some loss of power to detect QTL effects (Churchill & Doerge, 1994).

Detection of QTLs of a certain size is not guaranteed, but only carries a greater or lesser probability (Haley & Andersson, 1997). The probability of detection is referred to as the power of the experiment. Factors affecting the power include the population type, the sample size, the effect of the QTL (that is, whether it has an additive, dominant or imprinting effect), the size of the effect, the number of QTL controlling a trait, and the marker density. Some QTLs that have significant support may be false positives, and QTLs responsible for significant variation within and between populations can be missed if the breeds that are crossed are fixed for alleles with similar effects (that is, if the QTLs are not unique to one of the breeds). Therefore, the best method of confirming the presence of QTLs affecting a trait is by repeating experiments independently using different populations and, if possible, a different breed cross (Lander & Kruglyak, 1995).

1.7.2.4. Phenotype Requirements

When testing for the presence of QTLs, the different genotypes occurring at each marker are compared to see if there is a phenotypic difference associated with them. Thus, substantial variation between animals for the quantitative trait within the experimental population is essential.

Most quantitative traits are affected to some extent by environmental influences. Hence, anything that can be done to reduce background variation in the phenotype caused by effects of the environment will increase the likelihood of detecting QTLs. Replicating measures of a trait on the same individual is very valuable if the repeatability of the trait is low due to uncontrollable environmental fluctuations at the time of measurement. Environmental influences which affect the trait of interest during the experiment may be unavoidable, for example if an experiment has to be carried out over several seasons or locations. The sex of an animal also often has a big effect on the mean of a trait. Such influences can be allowed for in the analysis, by removing (and estimating) their effects simultaneously with detecting and estimating QTL effects (Haley & Andersson, 1997).

1.7.3. QTLs for Behavioural Traits

1.7.3.1. QTLs in Cattle

QTLs have been characterised for a number of economically important traits in cattle (and other livestock species) since the early 1990s, as comprehensive marker maps have become available. Approaches to map QTLs are now well established, in national populations as well as in experimental resource herds. Most of the mapping efforts have targeted conventional traits such as milk yield and composition or growth and carcass characteristics, but attention is now shifting increasingly towards less conventional traits such as meat quality, fertility or disease resistance (Georges, 2000).

A large number of studies have been conducted to scan the genome for QTL affecting milk production traits (such as milk yield, fat yield, fat percentage, protein yield and protein percentage), health traits (for example, veterinary treatments, clinical mastitis, udder health or somatic cell counts), growth and carcass traits (for example, conformation scores, carcass weight, meat tenderness and fat yield) and calving traits (for example, calving ease, dystocia and still births). Most have been carried out using national populations in different countries (Georges *et al.*, 1995; Grupe *et al.*, 1998; Zhang *et al.*, 1998; Ashwell & Tassell, 1999; Elo *et al.*, 1999; Heyen *et al.*, 1999; Schrooten *et al.*, 2000; Klungland *et al.*, 2001; Ashwell *et al.*, 2001; Olsen *et al.*, 2002; Kuhn *et al.*, 2003; Boichard *et al.*, 2003; Viitala *et al.*, 2003). A smaller number have used resource populations resulting from breed crosses (Stone *et al.*, 1999; Keele *et al.*, 1999; Casas *et al.*, 2000, 2001; MacNeil & Grosz, 2002). Many of these studies measured the same traits, and mapped QTLs to the same chromosomal regions independently, providing good confirmation of the presence of the QTLs.

1.7.3.2. Behaviour QTLs in Cattle

Only a few very recent studies have looked for QTLs influencing behavioural traits in cattle. Schrooten and colleagues (2000) included two 'workability' traits, 'milking speed' and 'temperament', in a genome scan for a large number of conformation and functional traits in Holstein Friesians in the Netherlands. They localised QTLs on three chromosomes for milking speed, but these were not significant at a genome-wide level, and no QTLs were localised for temperament. Two QTL studies have mapped measures of behaviour in resource herds. Six QTLs for temperament measures were found in the Canadian Beef Cattle Reference Herd (Schmutz *et al.*, 2000). Analysis is underway to confirm five putative QTLs for behaviour and stress responses in a Limousin x Jersey resource herd in New Zealand and Australia (Fisher *et al.*, 2001). These two studies are described in more detail in Chapter Six. As described above, the only way to truly confirm the location of a QTL acting on a trait is to repeat the finding in another population. Therefore more studies need to be carried out to provide strong evidence for the existence of the QTLs for the temperament traits identified in these studies.

1.7.3.3. Behaviour QTLs in Mice

Indications that it should be possible to find QTLs that influence behavioural traits in farm animals come from the large number of studies that have mapped genes controlling behavioural traits in mice. Mapping studies have identified QTLs affecting a wide range of behavioural traits in mice, including ethanol acceptance and sleep time, learning and memory, spatial abilities and emotionality (Johnson *et al.*, 1992; Nguyen & Gerlai, 2002; Milhaud *et al.*, 2002; de Groot *et al.*, 2001). The genetic basis of fearfulness traits in cattle is expected to be similar to that in mice due to conservation of homologous genes between species (Flint *et al.*, 1995).

A series of papers describing QTL mapping studies for behavioural measures taken from various tests of anxiety or fear in mice have been published over the last eight years. The mice are often tested in an open-field (OF) test and measures of activity (OFA) and defecation (OFD) are taken. High levels of activity in the OF test is interpreted as showing low levels of fear. Other commonly used measures are the number of transitions between rooms in light-dark (LD) conflict tests, the number of entries into the arms of elevated plus-mazes (EPM), and freezing behaviour in tests of conditioned (learned) fear responses (de Groot, 2002).

Flint *et al.* (1995) measured activity in an OF test, Y-maze and EPM in F_2 progeny from a cross between two inbred mouse strains which had been selected for high and low OFA scores. Three loci were found on chromosomes 1, 12 and 15, and the authors proposed these to be, at least in part, the genetic basis for anxiety. A number of studies have been published since then, each of which reports several QTLs for OFA and other test measures. Some of the QTLs have been replicated by one or more other studies, providing confirmation of their effect, and some are unique to one study. OF tests and a LD assay were carried out on a cross between two inbred strains that differed in OF behaviour (Gershenfeld *et al.*, 1997; Gershenfeld & Paul, 1997). They confirmed the QTL found by Flint *et al.* (1995) on chromosomes 1 and 15, and also mapped QTLs on chromosomes 10 and 19. A high-resolution approach focussed on chromosomes 1, 12 and 15 was taken by Talbot *et al.* (1999), and they mapped a QTL for OFA and OFD to a 0.8 cM interval on chromosome 1 and a second QTL found on chromosome 12. Turri *et al.* (2001a) replicated the initial experiment (Flint *et al.*, 1995) with the same cross but a different population, and found QTL on chromosomes 1, 4, 12, 15 and 18 for OFA. In another experiment mice from the same cross were tested in five different behavioural tests (Turri *et al.*, 2001b), and found QTLs on eight chromosomes, some of which were general to all the tests and some of which were test specific. QTLs on chromosomes 1, 4 and 15 were associated with more than one test.

These studies independently confirm QTLs involved in fear-related traits on chromosomes 1, 12 and 15. The QTLs on chromosomes 1 and 15 have been mapped with high resolution to small regions of chromosomes (Turri *et al.*, 1999; Talbot *et al.*, 1999). This success in finding strong evidence for the presence of genes influencing fear-related traits in mice provides good reason to believe it should also be possible to identify QTLs for similar traits in cattle.

1.8. Aims

This review has described some of the methods that can be used to measure fearfulness traits in cattle, and highlighted some of the issues that should be addressed when doing so. Evidence for the effects of genetics on fearfulness traits in cattle has been detailed, along with recently-developed methods that make it possible to locate genes influencing such traits.

The aim of this project was to investigate further fearfulness traits in cattle and identify QTLs for the traits. In light of the findings of the review, four specific aims were formulated:

1. To investigate whether repeatable, validated measures of fearfulness could be obtained from cattle, using a range of behavioural tests.

2. To investigate whether measures obtained from the tests supported the hypothesis that fearfulness is a multidimensional trait in cattle.

3. To investigate whether carrying out behavioural tests at a young age could be useful in predicting reactions to similar situations when older.

4. To identify QTLs that influence measures of fearfulness.

Chapter Two: Methods

2.1. Introduction

This chapter describes the population of cattle used in the study, the procedures of the behavioural tests carried out on them, and the schedules followed when carrying out the tests. It also described statistical methods applied in the analysis of the data obtained from the tests, and details how the genotyping data was obtained, and the methodology of the QTL mapping analysis.

2.2. Animals

The cattle studied were part of the Roslin Institute Bovine Genome Mapping Herd (RoBoGen), a resource herd established by breeding pure-bred Holstein cows (an extreme dairy breed) with Charolais bulls (a widely used beef breed). Four founder Charolais sires were crossed with 111 Holstein cows to produce 132 Charolais x Holstein F_1 heifers and 8 F_1 bulls. The F_1 offspring were then crossed with either each other to produce F_2 intercross calves, or backcrossed to Charolais bulls or additional Holstein cows, to produce the experimental second generation cross (see Fig. 2.1). One hundred and ninety-seven heifers and 181 bull calves from the second generation were born in the spring over three years, from 1999 - 2001 (called 'Year 1', 'Year 2' and 'Year 3' respectively; see Table 2.1). Where possible, the sires were allocated to the same mates in successive years, producing some full-sibling families.

These second generation animals, which showed substantial variation in a wide range of traits (the most obvious being colour and size), were the subjects of this study. For ease of description, the F_2 and backcross animals are referred to as F_2 animals, unless explicitly stated. The herd offered a unique opportunity for collecting behavioural data from a large number of genetically distinct animals that were reared, as far as possible, under the same management conditions. Each animal was identified by a unique number on its ear-tag.

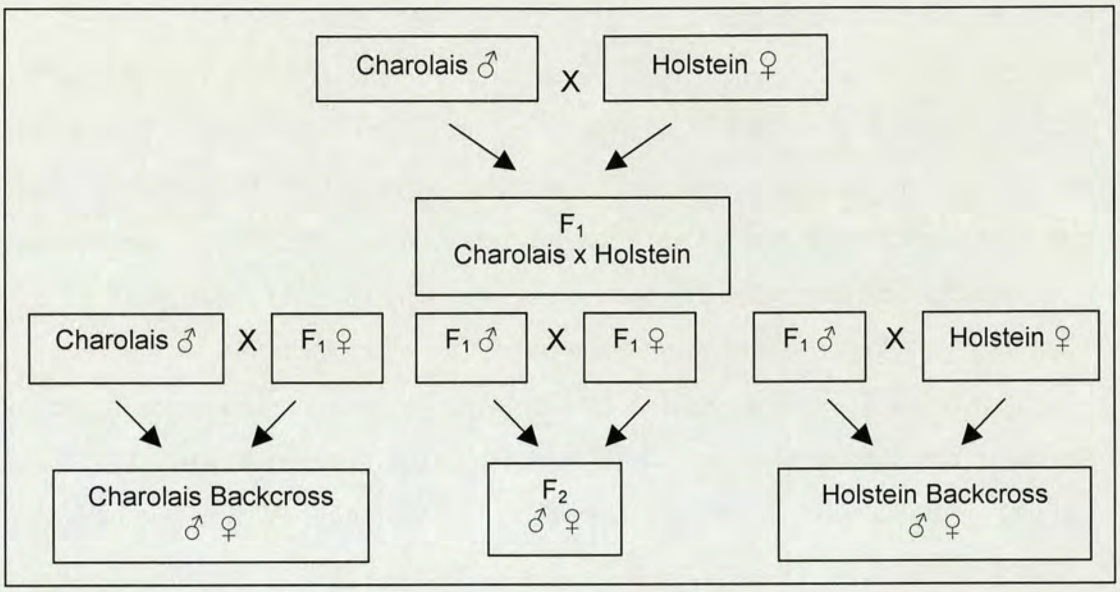


Figure 2.1: Breeding programme for the development of the RoBoGen herd. The second generation animals were the subjects of this study.

Table 2.1: Numbers of second generation experimental cattle.

Year of Birth	F2	Holstein Backcross	Charolais Backcross	Total Number
Bulls				
Year 1	28	4	17	49
Year 2	46	9	9	64
Year 3	52	4	12	68
Bull Totals	126	17	38	181
Heifers				
Year 1	23	18	15	56
Year 2	51	16	5	72
Year 3	48	10	11	69
Heifer Totals	122	44	31	197

Management procedures and standards for the second-generation calves were documented to ensure that the calves were reared in as similar environmental conditions as possible from year to year (Nicoll, 1999). Male and female calves were subject to different rearing conditions. Twenty-four hours after calving, F_2 heifer calves were removed from their dams and put in an individual calf pen to be artificially reared on a dairy calf bucket system. At two weeks of age they were housed in small groups, weaned off milk at approximately six weeks of age, and at 10 weeks moved into a large communal shed, where they were housed in groups of

11-14, and fed on concentrate and straw. The groups were composed of animals of similar weight (and therefore age). Their diet was designed so that the heifers reached breeding size and condition at approximately two years. Silage was introduced at 4 months. At one year of age they were put out to graze grass from May to October subject to conditions and grass availability. The F_2 heifers were inseminated when they were 21 months, and calved at two and a half years of age, when they were kept as dairy animals throughout their first lactation.

F_2 bull calves were suckled by their dams at pasture until weaning at six months of age. Backcross males calved from Holstein cows were fostered to F_1 cows that had produced heifer calves, where possible. At weaning they were also housed in a large communal shed in groups of 11-14, and received proprietary concentrate and roughage *ad libitum*. They were reared to a target slaughter weight of 550 kg, which was achieved at approximately 12 months of age. The animals were regularly handled for procedures such weighing and blood-sampling, which were undertaken as part of other experiments.

The behaviour tests (except the Dairy Parlour Scoring) were carried out while the animals were group-housed in the communal sheds. In each shed the pens were in a row and separated by gates, allowing contact between animals in adjacent pens, and allowing groups to be moved between the pens easily (Fig. 2.2, 2.3 & 2.4). An area of 6 m x 7 m could be closed off for testing in each pen in the heifer shed, and an area of 9.25 m x 8.5 m at the back of the pens in the bull shed. As part of another experiment, the feed intake of the animals was recorded using automatic feeders. As not all of the pens contained automatic feeders, the groups in some years were rotated round the pens at fortnightly intervals, to ensure all animals were on the automatic feeders for two weeks at a time.

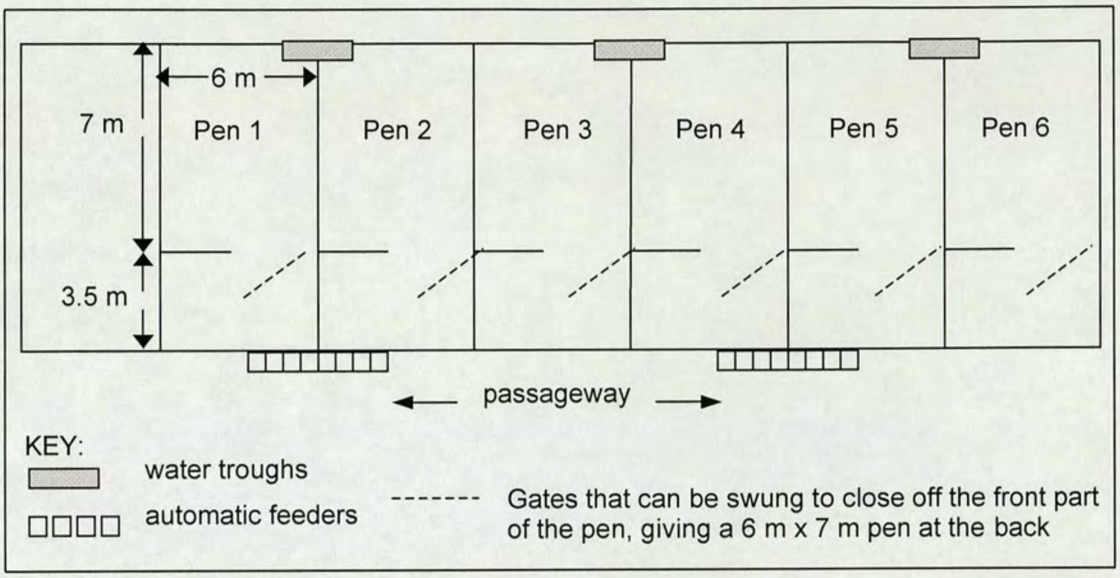


Figure 2.2: The layout of the pens in the shed where the heifers were housed and tested (not to scale).

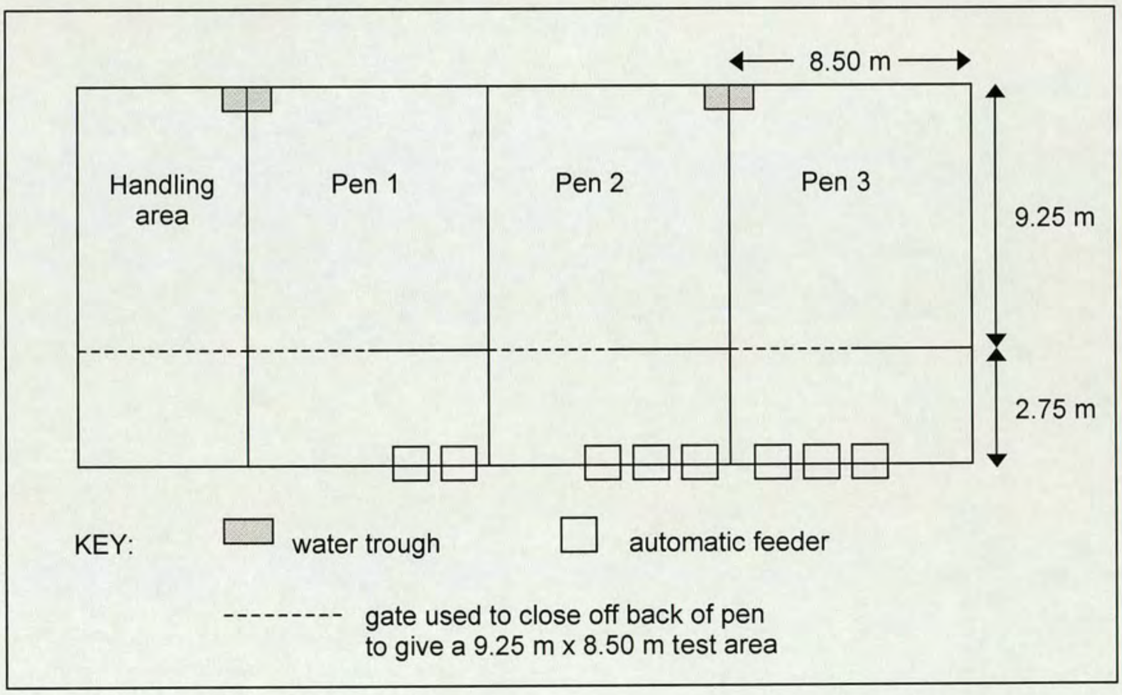


Figure 2.3: The layout of the pens in the shed where the bulls were housed and tested (not to scale).

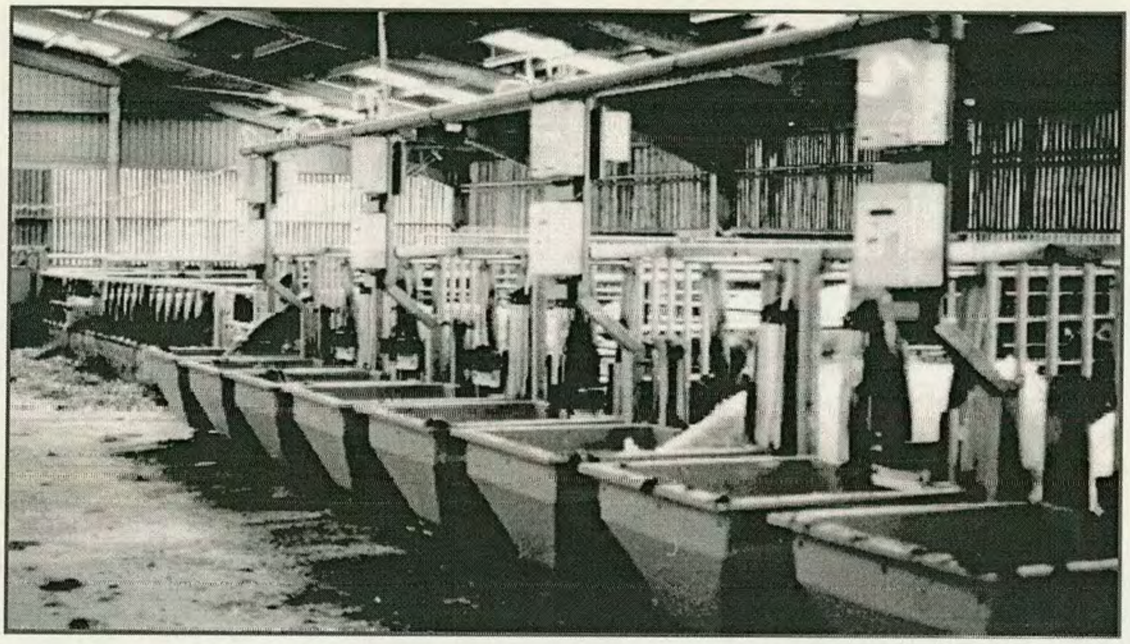


Figure 2.4: A photograph of the automatic feeders and pens in the shed where the heifers were housed and tested (photo courtesy of John Williams).

2.3. Behavioural Tests

2.3.1. Introduction

Four temperament tests were carried out on a large number of the F_2 animals to measure behavioural responses to particular stimuli: a Flight-from-Feeder (FF) Test, a Social Separation (SS) Test, a Novel Object (NO) Test and a Handling (HA) Test. Each test procedure and the analysis methods used for each are described in this section. The motivations thought to be underlying the behavioural reactions to each test situation, and the reasons these tests were chosen, have been discussed in Chapter One. These four test procedures were used in several experiments, described in Chapters Three to Six. Initially, each test was performed on a group of bulls and a group of heifers, to determine whether they gave results that were useful measures of temperament traits. This experiment is described in Chapter Three.

Further tests were also carried out on smaller groups of animals. A Sociality (SO) Test was carried out on a group of heifers in the final year of testing, as part of the experiment described in Chapter Four. Observations of behaviour in the dairy parlour were carried out on a group of F_2 heifers that calved in the final year of the project, and this experiment is described in Chapter Five. These two procedures are also described in this chapter.

2.3.2. Flight-from-Feeder Test

The FF Test was designed to measure the animals' levels of fearfulness of human approach. It was carried out while the animals were housed in pens with automatic feeders, as the animals were approached while at the feeders. The feeders were Insentec 'Hoko Farms' design, arranged in blocks of four per pen. The design of this test allowed each animal to be tested while remaining in its home pen with its penmates, yet the feeders had solid sides, preventing the test animal seeing and being affected by movements of the other animals around it. It was intended that this design should prevent the animals experiencing separation from their penmates during the experiment, so that fearfulness of humans was the only major motivation affecting the animals' responses (Fig. 2.5).

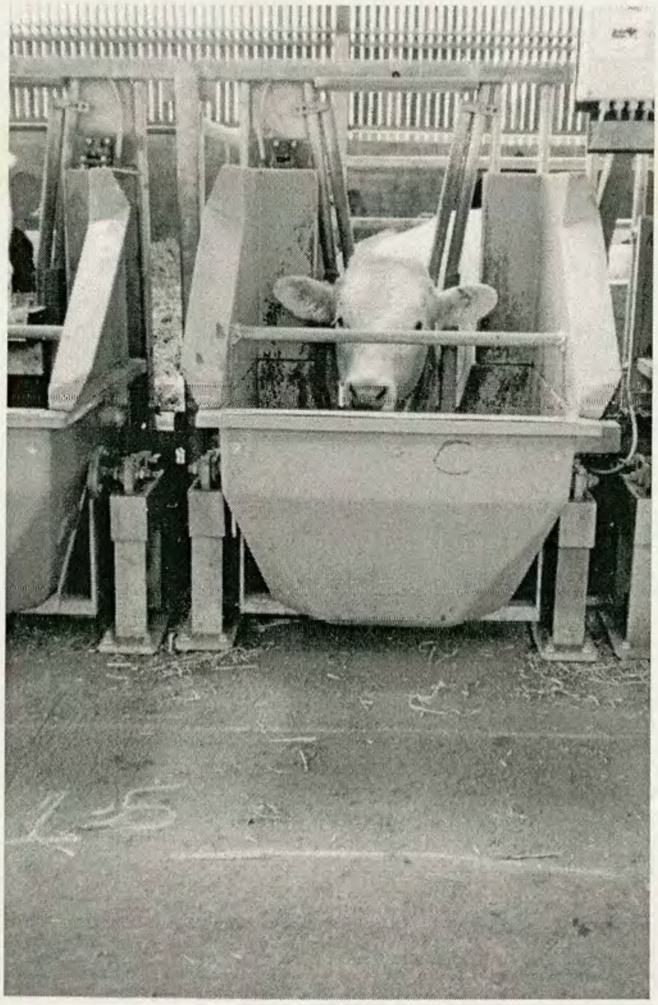


Figure 2.5: A photo of an animal feeding in an automatic feeder, with lines drawn on the floor ready for testing in the FF Test.

Chalk lines on the floor of the passageway in front of the pen marked the distance away from the pen to 2.0 m, at 0.5 m intervals (Fig. 2.6). The observer walked up and down the passageway in front of the pen at a distance of 2.5 m until an animal was spotted entering an automatic feeder. The observer then moved to a position 2 m directly in front of the feeder in the passageway, and stood stationary for 20 seconds while the animal fed, to determine a feeding bout was commencing. After 20 seconds the observer walked slowly towards the animal. If the animal backed out of the feeder as the person approached, the distance remaining between the observer and the feeder was recorded to the nearest 0.25 m. If the animal did not move away when the observer reached the feeder, the observer reached out and attempted to touch it on the forehead, and its response was recorded. The animal was then given a score for the test from the categories listed in Table 2.2. The score was noted on a score-sheet along with the ID number of the animal, date and time of day.

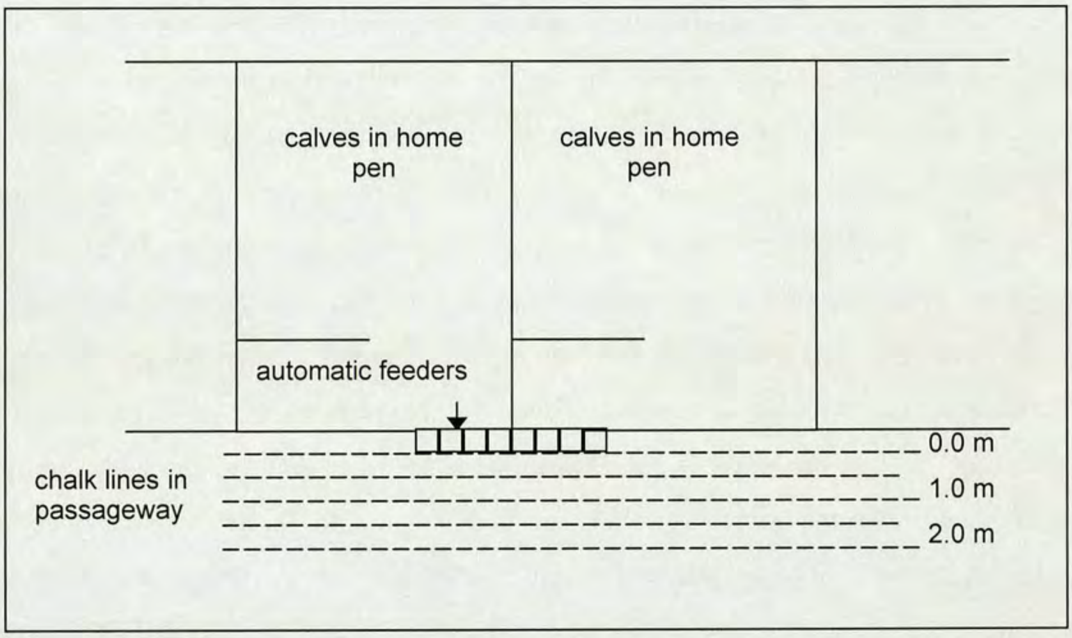


Figure 2.6: An example of the set-up for the Flight-from-Feeder Test for the heifer and bull calves (not to scale).

Repeated tests were carried out over the course of several days. Once tested, an animal was not approached again for at least 20 minutes, to avoid it being approached on consecutive visits to the feeders. During the test period the bulls had *ad libitum* access to pellet feed from their feeders, and so each was likely to have been experiencing similar low levels of hunger. The heifers were under restricted access to silage in the feeders at the time, and so were likely to have been experiencing higher hunger levels and feeding motivation than the bulls.

Table 2.2: The ordinal scale used to score the animals on their responses to the Flight-from-Feeder Test

Score	Response
0*	calf would not approach feeder while observer present
1	calf moved back when observer was 2.0 - 1.25 m away
2	calf moved back when observer was 1.0 - 0.25 m away
3	calf moved back when observer was 0 m away
4	calf moved back as observer extended arm to touch
5	calf moved back as observer touched it on the forehead
6	calf did not move back when touched on the forehead

* The 0 score was used only for Year 3 animals

2.3.3. Social Separation Test

The SS Test measured the animals' responses to separation from their penmates (sociality), and was also carried out in their home pen, so that novelty of the surroundings did not impact on the test. The group to be tested was moved two pens down the shed (Fig. 2.7) and the pens on either side of the test pen were cleared of animals. Plywood sheets were hung on the gates of the test pen to a height of 1.2 m. These sheets isolated the test animal from its pen-mates and other cattle, unless it made a deliberate attempt to look over the top of the boards or through gaps between the boards. The wooden sheets were hung in the pens a few days prior to the experiment so that the animals got used to them, in order that the animals were not also responding to the boards as novel objects during the SS Test.

After the group was allowed to settle for 10 minutes, each animal was moved into the test pen one at a time for five minutes. The test animal remained in auditory and olfactory contact with the rest of the group. The animals' responses were recorded on video-tape to be scored later.

A pilot trial suggested that animals that were separated from their pen-mates in this way resulted in sociality levels so high in some animals that they attempted to jump the gates of the pen. To reduce the extent of the isolation slightly, it was decided to put two 'companion' animals in the empty pen between the test group and the test pen. The companions were kept in the half of the pen furthest away from the test pen by two experimenters (Fig. 2.7). The test animal could see the companions by looking through gaps in or over the boards, but could not make physical contact with them. Any repeat tests were carried out on consecutive days.

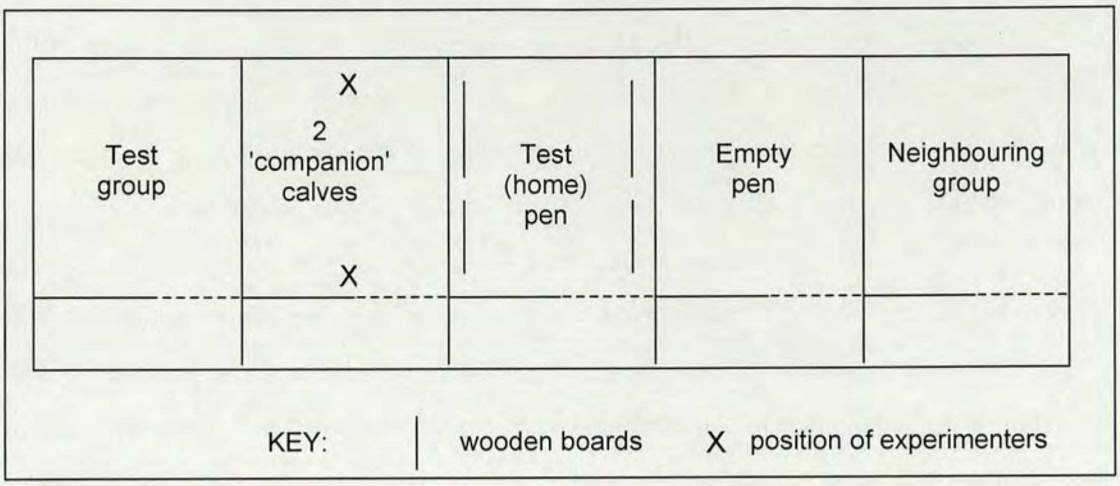


Figure 2.7: An example of the set-up for the Social Separation Test for the heifers (not to scale).

The video-tape of each test was observed, and continuous observations of the animals' behaviours over the five-minute test period were recorded using the Noldus 'Observer for Windows' computer package. Frequencies and total durations of the behaviours listed in the ethogram in Table 2.3 were calculated. Principal Components Analysis was used to group behavioural variables, as described in 2.5.4.

Table 2.3: The ethogram used for the Social Separation Test

Behavioural States and Events recorded in the SS Test	
States	
'Walking'	
'Running'	
'Gambolling'	(running with an arched back and tail held high, tossing head)
'Lying'	
'Kneeling'	(includes kneeling to rub head in straw or paw at straw)
'Escape'	(includes attempting to jump over gate or at back wall, standing with head over boards, and standing with head through bars at front of pen)
'Standing Alert'	(standing with head up and ears pricked)
'Standing Occupied'	(includes standing and: scratching or licking body, cudding or eating straw, or drinking; standing with nose in contact with bars at front of pen, wall at back of pen, or floor of pen)
'Interacting with Boards'	(standing with nose in contact with board or rubbing against board)
Events	
'Vocalise'	
'Urinate'	
'Defecate'	

2.3.4. Novel Object Test

The NO Test measured the animals' responses to a novel object placed in the home pen. The group to be tested was moved into the neighbouring pen, and allowed to settle for 10 minutes before testing began. Wooden boards were hung on the gates to prevent the animals seeing the novel object until it was their turn to be tested (see Fig. 2.8, 2.9). An orange traffic cone (base 44 cm wide, height 78 cm) or a green plastic barrel (diameter 55 cm, height 95 cm) was placed in the centre of the test pen. Each animal in turn was put into the pen with the novel object for five minutes. The test animal remained in auditory and olfactory contact with the rest of the group, and had contact with the usual neighbouring group which was present in the other neighbouring pen. Once the test animal had been put in the pen, the two experimenters stood at the far end of the neighbouring pen, so the animal could not see them during the test. The novel object was washed with water in between each test. The animals' reactions to the novel object were recorded on video-tape for later observation. Latencies, frequencies and durations of behaviours shown over the five-minute test were recorded from the videos using the Observer software and the ethogram listed in Table 2.4. The measures used in analysis were 'Latency to touch the novel object' and 'Time spent in contact with novel object'. The majority of the tests were conducted using a traffic cone, and the barrel was only used when repeated tests were carried out.

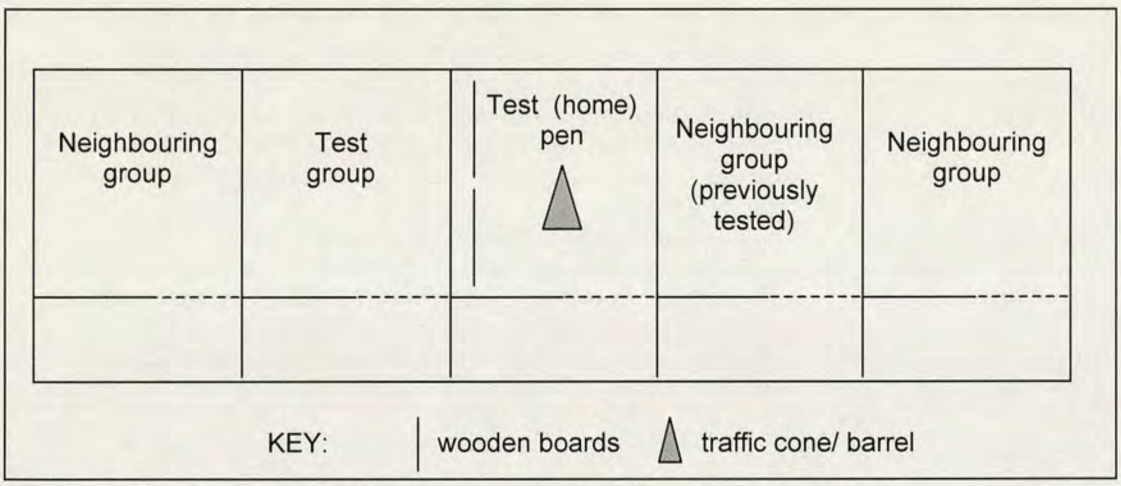


Figure 2.8: An example of the set-up for the Novel Object Test for the heifers (not to scale).

Figure 2.9: Photograph of a calf sniffing a traffic cone in the Novel Object Test.

Table 2.4: The ethogram used for the Novel Object Test

Behavioural States and Events recorded in the NO Test	
States	
'Walking'	
'Running'	
'Gambolling'	(running with an arched back and tail held high, tossing head)
'Lying'	
'Kneeling'	(includes kneeling to rub head in straw or paw at straw)
'Escape'	(includes attempting to jump over gate or at back wall, standing with head over boards, and standing with head through bars at front of pen or bars of neighbouring pen)
'Standing Alert'	(standing with head up and ears pricked)
'Standing Occupied'	(includes standing and: scratching or licking body, cudding or eating straw, or drinking; standing with nose in contact with bars at front of pen, wall at back of pen, or floor of pen)
'Interacting with Boards'	(standing with nose in contact with board or rubbing against board)
'Standing Near'	(standing with head stretched towards NO, *within* 1 headlength)
'Standing Far'	(standing with head stretched towards NO, *outside* 1 headlength)
'Standing Contact'	(standing with nose in contact with novel object)
'Standing Rubbing'	(standing or kneeling and pushing or rubbing against novel object with head or neck)
'Prancing'	(prance at cone playfully)
Events	
'Vocalise'	
'Urinate'	
'Defecate'	

2.3.5. Handling Test

This test was based on the Docility Test described by Boivin *et al.* (1992a,b) and Le Neindre *et al.* (1995). It was designed to measure the animals' responses to handling by humans. The test group was moved to the neighbouring pen on the side of the drinking trough (see Fig. 2.10), and the home pen was used as the test pen. No boards were used, so the test animal had contact with penmates in one neighbouring pen, and the familiar neighbouring group in the other. After moving, the test group was allowed to settle for 10 minutes before testing began. Two people were required for the test, a handler and a timer. The same person (myself) carried out the handling in all the tests.

There were three parts to this test:

i) Separating the test animal. The handler separated the test animal from the holding pen. The timer assisted at the gate, and recorded how long it took for the handler to bring the animal out into the test pen, and shut the gate behind it.

ii) Before handling. Once the animal had been put in the test pen, it was left alone for 30 seconds. Then the handler climbed into the corner of the pen and stood motionless for another 30 seconds.

iii) Handling. The handler attempted to contain the animal in a defined 2 m x 2 m area in a corner of the pen, for 30 consecutive seconds (see Fig. 2.11). The area was marked out by small chalk lines drawn on the wall and gate. If the handler succeeded in this, the handler tried to touch the animal on the shoulder, and the test ended. If the handler was unable to keep the animal in the corner after two minutes of trying, the test was ended. Steps ii and iii were recorded on video-tape.

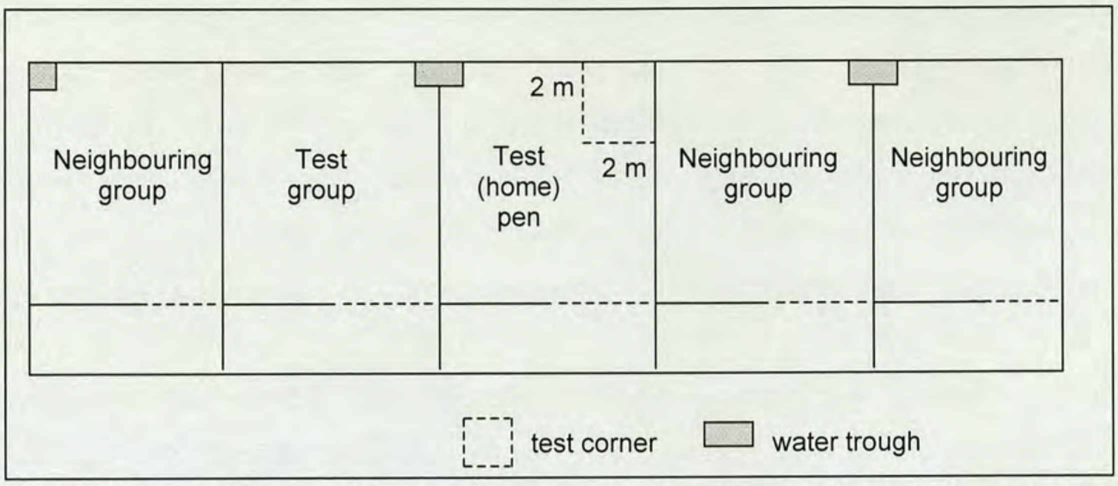


Figure 2.10: An example of the set-up of pens for the Handling Test for the heifers and the Year 1 bulls (not to scale).

Figure 2.11: A photo of a co-operative four-month-old heifer in the Handling Test.

Analysis of the behavioural data recorded on video during the test was carried out later using two methods. The first method again used the Observer computer package to calculate frequencies, latencies and durations of behaviours seen during the test. The ethogram used is shown in Table 2.5. The main measure taken was 'Latency to stand in the test corner for 30 seconds'. The second method used a computer program which was designed specifically for the analysis of the handling test, and kindly provided by Pierre Le Neindre (INRA, Saint Genès Champanelle, France). The program uses several variables that relate to the general activity and performance of the heifers during different stages of the test, and produces a Docility Score based on these (Le Neindre *et al.*, 1995). The variables used were; aggressiveness during different parts of the test, time spent in the corner, the number of times the animal tried to escape, time spent running, the interval(s) between the start of the test and standing in the corner for different lengths of time, and whether the animal could be stroked. Each of these variables was given a weighting, and their occurrence determined the Docility Score assigned to each animal. The method of calculation of the score is shown in Table 2.6.

Chapter 3.3.4.1 describes a preliminary comparison of these two methods. After this comparison, it was decided to analyse the data from the rest of the HA Tests using only the first method, that is, the method that used frequencies, latencies and durations calculated using Observer.

Table 2.5: The ethogram used for the Handling Test

Behavioural States and Events recorded in the HA Test	
States	
'Walking'	
'Running'	
'Gambolling'	(running with an arched back and tail held high, tossing head)
'Lying'	
'Kneeling'	(includes kneeling to rub head in straw or paw at straw)
'Escape'	(includes attempting to jump over gate or at back wall, standing with head through bars of pen)
'Standing Alert'	(standing with head up and ears pricked)
'Standing Occupied'	(includes standing and: scratching or licking body, cudding or eating straw, or drinking; standing with nose in contact with bars at front of pen, wall at back of pen, floor of pen or other animal)
'Time in Corner'	(time spent standing in the test corner)
Events	
'Vocalise'	
'Urinate'	
'Defecate'	

Table 2.6: Method used to calculate a Docility Score at the different steps of the Docility Test[a].

Variable	Value if X < Step 1	Value if X > Step 1 and < Step 2	Value if X > Step 2	Step 1	Step 2
Aggressiveness[1]					
When sorting the subject	Score=6.5				
Before handling	Score=7.5				
During first 30s handling	Score=8.5				
After first 30s handling	Score=9.0				
Time spent in corner (%)[2]	0	0.5	1.0	20	40
No. times animal tried to escape(%)[2]					
Before handling	0	-0.25	-0.5	1.0	≥1.0
During handling	0	-0.125	-0.25	≤0.3	≥1.7
Time spent running (%)[2]					
Before handling	0	-0.25	-0.5	2	8
During handling	0	-1.25	-0.25	5	15
Interval between beginning of the test and entrance in the corner for a time of less than[2]: 10s	0	-0.2	-0.4	30	60
15s	0	-0.2	-0.4	30	60
20s	0	-0.2	-0.4	30	60
25s	0	-0.2	-0.4	30	60
30s	0	-0.2	-0.4	20	40
Maintained for 30 consecutive seconds in the corner[2]	0	1.0	-	0.5	
Stroking of animal[2]	0	1.5	-	0.5	

[a]The Docility Score is calculated by summing 13.5 with either a set score if the animal shows aggressiveness during the test[1], or values of 'X' given for each of the other variables measured[2]. Each value of X depends on whether the animal scores higher or lower than limits set for Steps 1 and 2, (given in the last two columns). Adapted from Le Neindre *et al.*, 1995.

2.3.6. Sociality Test

The SO Test was designed to measure the animals' sociality (the extent to which they seek social companionship), by measuring their social reinstatement response when separated from the group. A group of animals was moved to a holding pen. In turn, each animal was herded to the far end of a corridor 36 m long, and then released and its behaviour recorded. The behaviour measures taken were based on those taken from quail in a runway test (Jones *et al.,* 2002).

The set-up is shown in Figure 2.12. The holding pen was familiar to the animals as they were brought there weekly for weighing. The corridor could be easily seen from the holding pen, but the animals were unlikely to have been in it before the first Sociality Test. Alongside the corridor is a row of calving pens, a few of which held dairy cows and calves while the test was being carried out. In order to block the animals in the calving pens from view, wooden boards were hung on the gates. The wooden boards had been hung in the animals' home pens prior to and during the SS Tests, and therefore were familiar to the animals.

A gate was swung across the corridor from Pen 2 to allow test animals to be held one at a time at the far end of the corridor (see Fig. 2.12). Chalk lines were drawn on the wall and on the boards at two places down the corridor, one of which defined the 'close zone'. Sawdust was scattered on the floor to make the surface less slippery.

One group at a time was moved to the holding pen and allowed to settle for 15 minutes. Each heifer was taken up to the start pen at the far end of the corridor, and allowed to settle for 30 seconds. The gate of the pen was then opened fully by Handler 1, and the heifer was then able to move up and down the corridor. The test ended three minutes after the gate of the start pen was opened, and then the test animal rejoined the group.

The behaviour of the heifers was recorded by video-camera, and latencies, counts and durations of behaviours were later scored from the videos using a stopwatch and recording sheet. The following measures were recorded: latency to reach 1st chalk line, latency to enter 'close' zone, and accumulated time spent in close zone. If the heifers did not cross either of the chalk lines, they were assigned a maximum latency of three minutes.

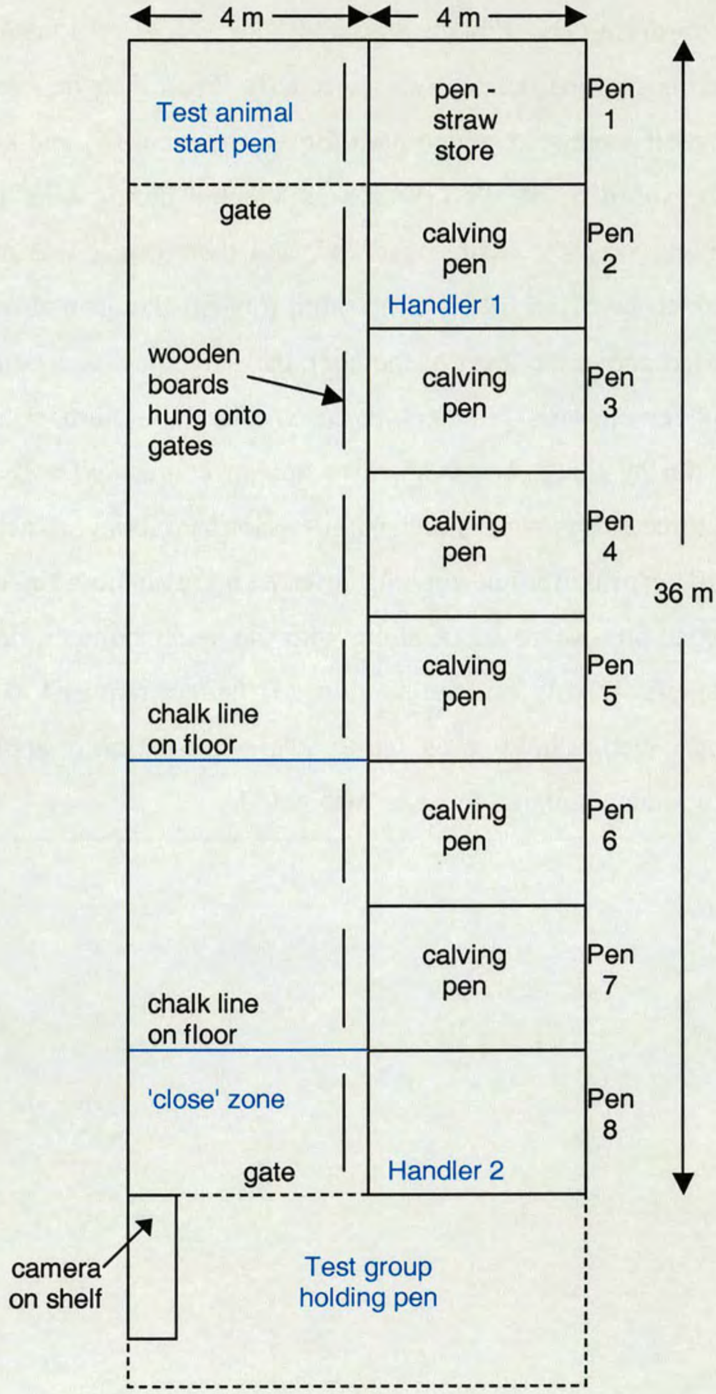


Figure 2.12: The set-up for the Sociality Test.

2.3.7. Dairy Parlour Behaviour Scoring

Behaviour shown by heifers while in close human contact when being milked was recorded by observing the 'Flinch, Step and Kick (FSK) response' reported by Willis (1983). During milking, some cows constantly flinch their udders or stomach muscles (F), shift their weight from one hind foot to the other (S) and kick (K). The recording procedure used by Willis (1983) was adapted during pilot trials, which showed that flinching was not easily observed, and therefore it was not recorded. Any movement which involved lifting either hind foot off the ground was recorded. If the foot was raised above the level of the hock the movement was counted as a K, and any other movement was counted as an S. The total number of S and K responses shown during three stages of close human contact were recorded from each animal. The three stages were udder-wiping prior to milking, attachment of the clusters at the start of milking, and dipping tests with antiseptic after milking. The counts were recorded on a score sheet, along with the heifer number, date, time and identify of the person carrying out the handling. If heifers required extra attention during these periods, such as holding the tail to quieten the heifer, or applying a kick-bar to a hind leg to reduce kicking, this was also noted.

2.4. Testing schedules

2.4.1. Introduction

The behaviour tests described in the last section were carried out on the F_2-generation animals born in Years 1 to 3. Each year, the animals were housed in groups of similar weight (and therefore approximately similar age), and each year these groups were named alphabetically in order of decreasing weight. The bulls were tested in the winter when they were approximately eight months of age. The heifers were tested in the spring when they were approximately 10 - 12 months of age. A schedule of tests, with a specified number of days between different tests, was followed within each year. Additional tests were carried out on the heifers born in Year 2 when they were four months of age. Some heifers born in Year 1 were also observed in the dairy parlour at the age of 30 months. The schedule used for this experiment is described fully in Chapter Five, and is not described here.

FF Tests were carried out on all the bulls and heifers each year, and SS Tests were carried out on all of the heifers each year, for the genetics experiment described in Chapter Six. The other tests were carried out on some of the animals in some of the years, as required for the objectives of the experiments described in Chapters Three to Five. Therefore the schedules varied between years. The following sections describe the test schedules that were followed each year. The animals tested and the tests carried out in each year are summarised in Table 2.7.

The FF Tests were carried out on the groups either *after* all the other tests had been carried out on all the groups (bulls), or *before* all the other tests (heifers). They were carried out all day, between the hours of 8:30 and 16:30, unless otherwise stated in the following sections. As the time interval between repeats of this test was dependent upon when the animals approached the feeders, the FF test was repeated on each animal either on the same day or up to eight days later.

SS, NO and HA Tests were usually carried out on two groups per day. There were between three to six groups of each sex per year. Each testing schedule was carried out on the first two groups over the course of one or two weeks, and then the same schedule was repeated on further pairs of groups until all the groups had been tested. Generally, the SS, NO and HA Tests were carried out between the hours of 11:15 and 17:30, unless otherwise stated, and repeats of the tests were usually carried out on consecutive days.

The SO Test was carried out on some of the heifers from Year 3 only. Contrary to the other tests, the SO Tests were carried out in the morning, between 9.30 and 13.30 (as the facility used was required by the dairy in the afternoon).

Table 2.7: Summary of the tests carried out on the F_2-generation animals.

Year	Animals	Age (months)	Tests	Number of Animals Tested
Year 1	Bulls	10	FF, HA, NO, SS	49
	Heifers	12	FF, HA, NO, SS	56
		30	Dairy Behaviour Scoring	32
Year 2	Bulls	10	FF, HA	64
	Heifers	4	HA, NO, SS	72
		12	FF, HA, NO, SS	72
Year 3	Bulls	10	FF, HA	68
	Heifers	12	FF, SS, (SO)	69 (42)

The tests used were: FF = Flight-from-Feeder Test, HA = Handling Test, NO = Novel Object Test, SS = Social Separation Test, SO = Sociality Test.

2.4.2. Year 1 Animals

2.4.2.1. Test schedules for the bulls

At the time of testing, 49 bull calves born in Spring 1999 (Year 1) were housed in four groups (A - D) with approximately 10 animals in Groups A, B and D, and 19 animals in Group C. The bull shed was still under construction at this time, and so the bulls born in Year 1 were housed in the same shed as the heifers at the time of testing (Fig. 2.2). Group C was kept in a double pen, in order that all the animals were housed in pens that allowed access to automatic feeders.

Four tests, the FF, SS, NO and HA Tests, were carried out on the bulls. The HA Test was carried out first, in December 1999, when the median age of the bulls was 244 days, or approximately eight months (range 164 - 251 days). Groups A, B and D were tested between 11:30 and 15:30, and Group C was tested between 09:15 and 11:15. The test was repeated three times on each group, with two non-test days between subsequent tests (see Table 2.8). The order of testing of the animals was selected randomly on the first day, and the same order was used in all subsequent tests.

The NO Test (using a traffic cone) and SS Test were carried out approximately one and a half - two months later. One NO Test and three SS Tests were carried out on four successive days on Groups A and B. This schedule was due to be repeated a week later with Groups C and D, but was delayed after the first SS Test on Group C and restarted the following week, due to a query concerning the Home Office License under which the animals were registered for other experiments, and was beyond the control of this experiment. The NO Tests and SS tests were all carried out between 12:00 and 17:45. The same order of testing of the animals was used as in the HA Tests.

Table 2.8: The test schedule for the 49 Year 1 bulls

Test	Group			
	A	B	C	D
Handling 1	Day 1[1]	Day 5	Day 6	Day 14
Handling 2	Day 5	Day 8	Day 9	Day 17
Handling 3	Day 8	Day 11	Day 12	Day 20
Novel Object	Day 61	Day 61	Day 68	Day 68
Social Separation 1	Day 62	Day 62	Day 69	Day 75
Social Separation 2	Day 63	Day 63	Day 75[2]	Day 76
Social Separation 3	Day 64	Day 64	Day 76	Day 77
Flight-from-Feeder	Day 91 - 94	Day 97 - 99	Day 91 - 99	Day 97 - 99

Day 1 was the first day of behaviour testing, 02/12/99

[1] An extra day was left in between the 1st and 2nd Handling Test of Group A

[2] Repeat testing of the Social Separation Test on Group C was delayed by 6 days.

Finally the FF tests were carried out approximately one month later again, when the median age of the bulls was 332 days, or approximately 11 months (range

249 - 343 days). The tests were carried out by an observer who had been present during the other tests and was therefore familiar to the animals. By then the groups of bulls were being moved to a new shed, so Group A and half of Group C were tested in the original shed (Fig. 2.2), and Groups B, D and half of Group C were tested in the new shed (Fig. 2.3). The arrangement of feeders was the same in both sheds, allowing the same test procedure to be carried out in each. Each animal was tested two or three times over the course of nine days, with most of the repeat tests on each animal carried out within three days. The animals were tested in whatever order they approached the feeders.

As these were the first tests carried out on any of the animals, they served to some extent as a pilot trial. The design of the schedules was altered slightly for testing the rest of the animals.

2.4.2.2. Test schedules for the heifers

The same four tests, the FF, SS, NO and HA were carried out on 56 heifers born in Year 1. The animals were tested from March to May 2000, when they were approximately 11 months of age. The heifers were housed in the heifer shed (see Fig. 2.2) and at this time were housed in three groups of approximately 18 animals. The tests were carried out in a different order from the bulls. The FF Tests were carried out first, when the observer was unknown to the calves. They took place when the median age of the heifers was 327 days, or approximately 11 months (range 274 - 360 days). Each animal was tested two or three times over the course of eight days. The animals were tested in whatever order they approached the feeders.

The other three tests were carried out just over one month later, by which time the heifers were housed in four groups (A, B, C and D) of approximately 13 animals per group. The tests were carried out in the afternoon, and two groups were tested per day. Each group followed a ten-day testing regime as detailed in Table 2.9. For example, three SS Tests were carried out on Group A on Days 41, 42 and 43, a NO Test on Day 45 and three HA Tests on Days 48, 49 and 50. No tests were carried out on the intervening days. A traffic cone was used as the novel object in the NO Tests. The median age of the heifers was 367 days or approximately 12 months (range 319 - 405 days) on the day of the first SS Test. The order the animals were tested in was selected randomly for the first SS Test, and the same testing order was used in all subsequent tests.

The tests were carried out in a different order from the bulls to attempt to minimise the extent to which responses were affected by factors other than those that the tests were designed to measure. For example, the SS Test was carried out before the NO and HA Tests as it was thought the unavoidable element of separation from the herd involved in the latter two tests would seem smaller if the animals had already experienced the much larger degree of separation in the SS Test.

Table 2.9: The Test Schedule for the 56 Year 1 heifers

Test	Group			
	A	D	B	C
Flight-from-Feeder*	Day 1 - 12	n/a	Day 1 - 7	Day 10 - 12
Social Separation 1	Day 41	Day 41	Day 55	Day 55
Social Separation 2	Day 42	Day 42	Day 56	Day 56
Social Separation 3	Day 43	Day 43	Day 57	Day 57
Novel Object	Day 45	Day 45	Day 59	Day 59
Handling 1	Day 48	Day 48	Day 62	Day 62
Handling 2	Day 49	Day 49	Day 63	Day 63
Handling 3	Day 50	Day 50	Day 64	Day 64

* The heifers were in three groups at this time
Day 1 was the first day of behaviour testing, 15/03/2000

2.4.3. Year 2 Animals

2.4.3.1. Test schedules for the bulls

The 64 bulls born in Year 2 were kept in 4 groups (E, F, G and H) and were tested while housed in the bull shed shown in Figure 2.3. As there were only three pens in this shed, the groups were rotated fortnightly between pens in the bull shed and pens in a neighbouring shed. All the tests were carried out in the bull shed at least six days after the groups swapped pens.

The HA Test was carried out three times in December 2000, when the median age of the bulls was 235 days, or approximately eight months (range 182 - 265 days).

This year the tests were carried out on consecutive days, as they were on the Year 1 heifers. The order of testing of the animals in each group was chosen randomly on each day. Approximately 50 days after the HA Tests, two FF Tests were carried out over the course of two days. This schedule is summarised in Table 2.10.

Table 2.10: The Test schedule for the 64 Year 2 bulls

Test	Group			
	E	F	G	H
Handling Test 1	Day 1	Day 1	Day 10	Day 10
Handling Test 2	Day 2	Day 2	Day 11	Day 11
Handling Test 3	Day 3	Day 3	Day 12	Day 12
Flight Test 1 + 2	Day 53, 54	Day 53, 54	Day 67, 68	Day 67, 68

Day 1 was the first day of behaviour testing, 04/12/2000

2.4.3.2. Test schedules for the heifers

The 72 heifers born in Year 2 were tested at two ages, first at four months, and later at twelve months of age. For both sets of tests, the heifers were housed in the pens in the heifer shed shown in Figure 2.2.

At four months, they were housed in five groups of approximately 14 animals. SS, NO and HA Tests were carried out starting in August 2000, according to the schedule shown in Table 2.11. The testing schedule for these tests was the same as that carried out on the Year 1 heifers. For example, three SS Tests were carried out on Groups A and B on Days 1, 2 and 3, a NO Test on Day 5 and three HA Tests on Days 8, 9 and 10. No tests were carried out on the intervening days. The median age of the heifers on their first SS Test was 119 days, approximately four months (range 98 - 138 days).

Table 2.11: The Test Schedule for the 72 Year 2 heifers at four months of age

Test	Housing Group				
	A	B	C	D	E
Social Separation 1	Day 1	Day 1	Day 13	Day 13	Day 35
Social Separation 2	Day 2	Day 2	Day 14	Day 14	Day 36
Social Separation 3	Day 3	Day 3	Day 15	Day 15	Day 37
Novel Object	Day 5	Day 5	Day 17	Day 17	Day 39
Handling 1	Day 8	Day 8	Day 20	Day 20	Day 43
Handling 2	Day 9	Day 9	Day 21	Day 21	Day 44
Handling 3	Day 10	Day 10	Day 22	Day 22	Day 45

Day 1 was 09/08/2000

The tests at 12 months were carried out following a similar schedule to the one used at four months. The animals were housed in the same shed in six groups (E, F, G, H, I and J) with approximately 12 animals in each. Restriction on access to the animals due to the 2001 outbreak of Foot-and-Mouth Disease meant that testing began a month later than planned this year. FF Testing began in April 2001 when the median age of the heifers was 367 days, approximately 12 months (range 293 - 406 days). Access restrictions prevented all of the FF Tests being carried out before the other tests began, and some of them were fitted in around other test days (see Table 2.12). At the time of testing, the groups were being moved around the pens in a six-week rotation as part of another trial, as there are automatic feeders in only four of the pens. This also restricted the time when FF Tests could be carried out on each group. Therefore, the FF Tests were carried out on all the animals over the course of 30 days in total. All repeats of the test on the same animals were carried out within three days, except on two animals, which were repeat tested within 10 days. The tests were carried out between 8:30 and 19:45.

After the FF Tests, two SS Tests were carried out on successive days, then one non-test day was followed by two NO Tests on successive days, and finally two more non-test days were followed by two HA Tests on successive days (see Table 2.12). A traffic cone was used as the novel object in the NO Tests. The order of testing of the animals in each group was chosen randomly on each day.

Table 2.12: The Test Schedule for the 72 Year 2 heifers at 12 months of age.

Test	Group					
	E	F	G	J	H	I
FF 1 + 2	Day 1, 2, 22, 30	Day 3, 4, 5,	Day 22, 25, 26, 30	Day 3, 4, 5	Day 2, 3	Day 3, 4, 22
SS 1	Day 8	Day 8	Day 22	Day 22	Day 36	Day 36
SS 2	Day 9	Day 9	Day 23	Day 23	Day 37	Day 37
NO 1	Day 11	Day 11	Day 25	Day 25	Day 39	Day 39
NO 2	Day 12	Day 12	Day 26	Day 26	Day 40	Day 40
HA 1	Day 15	Day 15	Day 29	Day 29	Day 43	Day 43
HA 2	Day 16	Day 16	Day 30	Day 30	Day 44	Day 44

Day 1 was the first day of behaviour testing, 09/04/2001
The abbreviations for the tests used are: FF = Flight-from-Feeder Test, HA = Handling Test, NO = Novel Object Test, SS = Social Separation Test.

2.4.4. Year 3 Animals

2.4.4.1. Test schedules for the bulls

The 68 bulls born in Year 3 were housed in five groups (I, J, K, L and M) of approximately 13 animals in each. Similarly to the Year 2 bulls, the groups were again rotated between two sheds on a fortnightly basis. Tests were carried out in the bull shed five days after the animals were moved into the pens (see Fig. 2.3). The same schedule that was used for the Year 2 bulls was followed for the Year 3 animals. The HA Test was carried out three times on consecutive days, beginning at the end of November 2001. The median age of the bulls on the day of their first HA Test was 231 days, or approximately eight months (range 161 - 258 days). The order of testing of the animals was again chosen randomly each day. Approximately 50 days later, two FF Tests were carried out over the course of between two and nine days for each group. This schedule is summarised in Table 2.13.

Table 2.13: The Test schedule for the 68 Year 3 bulls

Test	Group				
	L	M	I	J	K
Handling Test 1	Day 1	Day 1	Day 13	Day 13	Day 13
Handling Test 2	Day 2	Day 2	Day 14	Day 14	Day 14
Handling Test 3	Day 3	Day 3	Day 15	Day 15	Day 15
Flight Test 1 + 2	Days 56, 58, 59	Days 56, 58, 59, 64	Days 70, 71	Days 70, 71	Days 70, 71

Day 1 was the first day of behaviour testing, 28/11/01

2.4.4.2. Test schedules for the heifers

FF and SS Tests were carried out on the 69 Year 3 heifers, and additionally, SO Tests were carried out on 42 of the heifers. The tests were brought forward by a couple of months and started in February, because at the time there were doubts about whether funding to keep the animals on the farm could be continued.

At the time of testing, the heifers were housed in pens in the heifer shed (Fig. 2.2), in six groups with 10 - 13 animals in each group. The six groups were labelled K, L, M, N, O and P. The median age of the heifers on the first day of testing was 309 days, or approximately 10 months (range 251 - 338 days). At the time of testing, the groups were again being moved around the pens in a six-week rotation as part of another trial, as there are automatic feeders in only four of the pens. This affected the order that the FF Tests were carried out on the different groups, as the test makes use of the automatic feeders. It also affected the order of testing of the groups in SS and SO Tests, as groups were not tested while housed in Pen 6 at the end of the shed (because neighbouring groups could not be held on both sides of that pen).

The test procedures were carried out according to the schedule shown in Table 2.14. The FF Tests were carried out first over the course of four days, with between one to four days spent on each group. The SS and SO Tests were then carried out on two groups at a time. Two repeats of the SS Test were carried out on consecutive days, and five days later, SO tests were carried out on three consecutive days. The SO Test was carried out on the first four groups to be tested only, which were K, L, N and P.

Table 2.14: The Test Schedule for the 69 Year 3 heifers.

Test	Group					
	K	L	N	P	M	O
FF 1+2	Day 1, 4	Day 11	Day 1, 3, 4	Day 1, 3	Day 11	Day 1, 3, 4
SS 1	Day 38	Day 38	Day 52	Day 52	Day 66	Day 66
SS 2	Day 39	Day 39	Day 53	Day 53	Day 67	Day 67
SO 1	Day 44	Day 44	Day 58	Day 58	X	X
SO 2	Day 45	Day 45	Day 59	Day 59	X	X
SO 3	Day 46	Day 46	Day 60	Day 60	X	X

Day 1 was the first day of behaviour testing, 19/02/02

The abbreviations for the tests used are: FF = Flight-from-Feeder Test, NO = Novel Object Test, SS = Social Separation Test, and SO = Sociality Test.

2.5. Statistical Methods

2.5.1. Introduction

Frequencies, total durations and latencies were the measures of behaviour taken from most of the tests, except the FF Test and Dairy Parlour Scoring, which produced categorical scores and counts respectively. For analysis of the Handling Test, in addition to the behavioural measures, a 'Docility Score' was calculated (see Section 2.3.5). Much of the analysis of the data involved standard biological statistical tests for parametric and non-parametric data, such as paired t-tests and correlation coefficients (Fowler & Cohen, 1990; Martin & Bateson, 1993). Where the distribution of data was not normal, attempts at transformation were made before non-parametric tests were used. Some of the more complicated or unusual statistical methods used are described below. The calculation of repeatability coefficients, the interpretation of correlation coefficients, and the principles of using Principal Components Analysis to analyse groupings of behaviours, are discussed.

2.5.2. Repeatability and REML

Each of the behavioural tests was carried out a number of times on the same groups of animals. This enabled the 'repeatability' of the scores from each test to be calculated. Repeatability indicates whether behaviours measured in a test represent consistent traits of the animals, or are merely momentary responses to particular situations.

When repeated measures of a behaviour are made on each individual, the phenotypic variance seen can be partitioned into variance *within* individuals (measuring the differences between the performance of the same individual) and variance *between* individuals (measuring the permanent differences between individuals; Falconer, 1989). The ratio between the within-individual and between-individual variance components is used to calculate the repeatability.

The Restricted Maximum Likelihood (REML; Patterson & Thompson, 1971) procedure in the statistical computer package GenStat for Windows (Release 4.21, Rothamsted Experimental Station, 2001) was used to obtain information on the sources and sizes of variability in the data sets. REML copes with unbalanced data that has a hierarchical structure. A linear mixed model was used to estimate the treatment effects and variance components. 'Animal identity' and 'test repeat number' were fitted as random effects in the model.

When the variance components had been obtained, the repeatability was calculated using the formula:

$$\text{repeatability} = \frac{\sigma_a^{2}}{\sigma_a^{2} + \sigma_r^{2}}$$

where σ_a^2 is the variance component associated with the animal, and σ_r^2 is the variance component associated with the repeat scores.

The value obtained indicates how consistent the animals' responses are. This is crucial when assessing the usefulness of a behavioural test (see Chapter 1.5.1.1). A low value indicates that the within-individual component varies widely between repeat tests, and that the animals are not reacting in a consistent manner each time. A repeatability value of ≥ 0.5 indicates that at least 50% of the variation seen between tests can be attributed to consistent individual responses. This value was therefore taken as a guideline as to whether tests were considered to give sufficiently repeatable results. Values calculated on data that was not normally distributed were interpreted with care, as they will be only approximate values.

REML was also used to investigate whether factors other than consistent differences between individual animals caused variation in test scores. The factors investigated were 'year', 'housing group', test order', 'age', 'sire' and 'breed group', and these are explained further in Chapter 6.3. The factors were fitted as random factors in linear models. In each model, 'animal identity' was used as the residual term. Variance components for each term were obtained. Any component that was twice as large as it's standard error was considered to have an effect on the test scores.

In all the REML analyses, both for repeatability and investigation of factors affecting test scores, the statistical significance of all the components investigated were calculated. The deviance of each model was obtained from the REML output. Then, each component was sequentially omitted from the model and the deviance was obtained before they were reinstated (Genstat 5 Release 3 Reference Manual, 1993, p. 571). The difference between the two deviances was compared with a chi-squared distribution with one degree of freedom, to assess the statistical significance of the difference due to each component.

2.5.3. Correlation Coefficients

Correlation coefficients give an indication of the association between two sets of measurements, and were used to look for relationships between measures taken on the same animals in different tests. Product Moment (r) or Spearman Rank (r_s) Correlation Coefficients were used depending on whether the data was parametric or non-parametric respectively (Fowler & Cohen, 1990). Correlations were calculated using the statistical software package Minitab Release 13.32. Interpretations of correlations were made as described in Table 2.15. A correlation of at least 0.7 was regarded as a meaningful relationship, as this indicates that approximately 50% of the variation in one measure is accounted for statistically by variation in the other measure (Martin & Bateson, 1993). Spearman rank correlations were sometimes used to double-check evidence of relationships between data from repeated tests, when repeatability calculations were carried out on data that was not normally distributed.

Table 2.15: Informal interpretations for statistically significant correlations (Martin & Bateson, 1993)

Value of coefficient (+ve or -ve)	Meaning
0.00 to 0.19	slight; almost negligible relationship
0.20 to 0.39	low correlation; definite but small relationship
0.40 to 0.69	moderate correlation; substantial relationship
0.70 to 0.89	high correlation; marked relationship
0.90 to 1.00	a very high correlation; very dependable relationship

2.5.4. Principal Components Analysis (PCA)

Principal Components Analysis (PCA) was used to analyse groupings of the behavioural variables obtained from the SS Test. The purpose of PCA is to explain as much of the total variation in the data as possible using as few factors (or 'principal components') as possible (Kleinbauen *et al.*, 1988). This is done by taking behavioural variables which are correlated, and finding combinations of them which produce a few uncorrelated principal components that measure different 'dimensions' in the data (Manley, 1986). It is assumed that the uncorrelated component scores are more meaningful than original variables, whose responses are influenced by one another (Frey & Pimentel, 1978).

Firstly, the variables were standardised, by calculating the correlation matrix of the variables. The variances of the principal components (the 'eigenvalues' of the sample covariance matrix) and their corresponding coefficients ('eigenvectors') were then calculated from the correlation matrix, using the statistical computer programme GenStat for Windows. The percentage variation explained by each principal component was also calculated, and examination of these in a scree plot indicated how many of the components were needed to explain the majority of the variation in the data. The eigenvectors of these components were plotted to examine which variables loaded the most strongly. The value of the loading represents the degree to which the parameter influences the component. Values that are close together in a diagram are highly correlated, and may have a common motivational background (Forkman *et al.*, 1995). Further analysis investigating correlations between variables and components, and the percentage variance of each variable involved in each component, was carried out if appropriate.

2.6. QTL Analysis

2.6.1. Introduction

The interval mapping of QTLs for traits of interest within a population was briefly described in Chapter 1.5. The tools required for such an analysis are: a population of three generations, a map of marker loci that covers the whole genome, genotyping data from these markers for all the animals, variation in the quantitative phenotype measured from the F_2-generation animals, and QTL mapping software.

The population structure has been described in Section 2.2, and the recording of the phenotypic data using behavioural tests was described in Section 2.3. The genotyping of this population was carried out as part of the parent project, and the results were provided for the analysis in Chapter Six. The marker map used and the procedure for recording the genotyping data from the markers are described in 2.6.2. 'QTL Express' software was used to carry out the QTL analysis, and this is described in 2.6.3.

2.6.2. Genotyping Data

The marker and genotyping data for use in the analysis reported in Chapter Six was provided from the parent QTL mapping project. As described in Chapter 1.7.1, the bovine genome map now contains over 1600 microsatellite DNA markers (Barendse & Fries, 1999). The bovine genome is estimated to be around 3000 cM. The target for the genotyping panel was to have one marker every 20 cM, and this therefore requires about 150 markers. This is a suitable density of markers for conducting an initial whole-genome scan to localise QTLs (Barendse & Fries, 1999).

Four hundred and fifty markers were selected from two publicly available genetic linkage maps, the International Bovine Reference Panel (IBRP; Barendse *et al.*, 1993, 1994, 1997), and the US Department of Agriculture (USDA) Meat Animal Research Center (MARC) Map (Bishop *et al.*, 1994; Kappes *et al.*, 1997). These maps are available on the WWW at http://www.cgd.csiro.au/ and http://sol.marc.usda.gov respectively. As some markers appear on one map and not the other, the positions of some of the markers had to be estimated. The markers were selected on the basis of their position in the genome. They were tested in the F_0 generation animals for ease of use in the lab. (amplification by PCR and scoring of allele sizes) and for their predicted information content. Markers that could track chromosomal inheritance in more than 50% of the matings were selected. A panel of one hundred and eighty two markers that were sufficiently polymorphic in this population of animals was chosen for use.

DNA from all the grandparent, parent and offspring animals was obtained from blood samples, which underwent standard phenol extraction. The DNA was sent to Geneseek Inc. (Lincoln, Nebraska, USA) for genotyping in yearly batches. The DNA was amplified using PCR. The published conditions were modified as necessary to optimise performance in the laboratory. PCR amplifications were loaded onto an ABI automated DNA fragment analyser and run on acrylamide gels. Allele sizes were scored using ABI Genescan 2™ and Genotyper 1™ software. The genotype data was sent back to Roslin electronically in batches for checking and entering into a project database. Each batch consisted of a number of files, each containing a list of animal numbers with two alleles sizes for each marker locus.

2.6.3. QTL Express

'QTL Express' software was used for the analysis of the data (Seaton *et al.*, 2002). It is software for QTL mapping with a web-based user interface, and uses regression analysis of data from an F_2 population derived from a cross between two inbred or outbred populations. A recent addition to the application, which is currently under trial, also allows the combined analysis of F_2 and backcross data.

There were two steps to the mapping analysis. Firstly, the probabilities of each individual from the F_2 generation inheriting each of the possible genotypes at specific chromosomal locations was calculated from the marker data. Secondly, these probabilities were fitted to the phenotypic observations using a least squares framework, and the test statistic (F ratio) was calculated. The locations that maximised the F ratio on each chromosome were the best estimates for the locations of QTLs. Additional fixed effects that explained trait variation were also fitted into the model.

The calculation of a significance threshold was then carried out, to indicate how strong the evidence was for the presence of QTLs at the estimated locations. Permutation tests were used to set significance thresholds at the single-point or chromosome-wide level (Churchill & Doerge, 1994). The different levels that the threshold can be set at allows for a trade-off between using too stringent and too lax a standard (Lander & Kruglyak, 1995). A single-position threshold estimates separately for each analysis point, of which there are many on each chromosome. If these are used, significant results will be obtained by chance (type 1 error). A chromosome-wide threshold provides an overall critical value that is valid for all points on a chromosome, and is higher than the single-position threshold. It is used to detect the presence of a QTL somewhere on the chromosome, while controlling the overall type 1 error rate (Churchill & Doerge, 1994). Significance thresholds were calculated at both levels.

This chapter has described various methods used in the collection, analysis and interpretation of the data obtained during this project, and will be referred back to frequently in the following experimental chapters.

Chapter Three: Variation and Repeatability of Behavioural Test Results

3.1. Introduction

The aim of this experiment was to determine whether it is possible to obtain reliable measures of fearfulness traits in cattle using behavioural tests. As discussed in Chapter 1.5.1, if measures of behaviour obtained from tests are to be regarded as reflecting underlying temperament traits, two criteria must be met: the measures must be demonstrated to be consistent across repeated tests, and they must be validated by demonstrating consistency in different situations. Additionally, the tests must be designed so that they are of the appropriate severity to evoke a range of responses from the animals tested, so that variation in response between individuals is seen.

Many studies of temperament in cattle have been carried out, but few have repeated tests on the same animals more than once to assess the reliability of the behavioural data obtained, or compared animals' reactions in different tests. Four behavioural tests that are commonly used to measure fearfulness of cattle to different frightening stimuli, flight tests, the Docility Test, novel object tests and open-field tests, were described in Chapter 1.5.2. In the experiment described here, four tests based on these were assessed for two of the criteria, variation in response between individuals, and repeatability of response when tests were carried out a number of times on the same animals. Validation of the tests will be examined in Chapter Four.

Two of the tests carried out, the Flight-from-Feeder (FF) Test, which was an adapted flight test, and the Handling (HA) Test, which was based on the Docility Test, were thought to measure fearfulness of humans. A Novel Object (NO) Test was carried out which was thought to measure fearfulness of unfamiliar stimuli (neophobia). The fourth test was a Social Separation (SS) Test that was based on the open-field test but carried out in a familiar environment. It was designed to measure the animals' fearfulness of isolation (sociality).

Although these types of tests have been carried out frequently on cattle, the results have seldom been examined on these criteria. Flight Tests have been used extensively in cattle, as described in Chapter 1.5.2.1. Most studies that have included measures of flight distance have been concerned with differences between groups of cattle of different breeds or treatments (Fordyce *et al.*, 1982; Murphey *et al.*, 1980; 1981; Kabuga & Appiah, 1992). Intra-herd differences have been noted in intensively-reared Holstein dairy cows (Purcell *et al.*, 1988; Albright, 1993), and in a Limousine x Jersey herd (Fisher *et al.*, 2000). Despite their widespread use, flight tests have been assessed for intra-animal repeatability only once, very recently. Fisher *et al.* (2000) tested Limousin x Jersey cross animals, some of which were reared artificially as dairy calves, and some of which ran with their dams until six months of age. Each animal was placed next to a pen of six or more of its penmates in a yard, and approached by an observer from a distance. The test was repeated three times at monthly intervals, and showed a high repeatability of 0.51 ± 0.03. Purcell *et al.* (1988) also repeated a flight test three times on their study animals, but results concerning repeatability were not presented or discussed.

The Docility Test has been used to measure variation between individuals in artificially reared Tarine, Montbeliarde and Friesian heifers (Boivin *et al.*, 1992a) and Limousin heifers reared in range systems (Le Neindre *et al.*, 1995). Both these studies reported variation between the animals. Two very recent studies reported correlation coefficients between the scores obtained from repeated tests. (It should be noted that a correlation coefficient of 0.7 is approximately equivalent to a repeatability coefficient of 0.5; see Chapter 2.5.2 and 2.5.3). A moderate correlation ($r = 0.6$) was found between scores obtained in repeated tests carried out two days apart on 38 10-month Salers and Limousin calves (Grignard *et al.*, 2000). The test was also carried out twice on a large group of Limousin heifers, at nine months of age and again seven months later, and again a moderate correlation was found ($r = 0.56$; Grignard *et al.*, 2001).

The studies that have used novel object tests as a stressful test condition to impose on cattle have all been concerned with differences between treatment groups, rather than differences between individuals. However, Boissy *et al.* (1998) noted that large variation was seen between individual Aubrac heifers. No attempts have been made to test the repeatability of novel object tests. That the animals have not encountered the novel object before is integral to the test. Therefore, repeats of the test would have to be carried out using a different novel object each time, and even then, animals may habituate to repeated presentation.

Inter-individual variation in behavioural responses to social separation tests has been demonstrated in dairy cows (Hopster & Blokhuis 1993), dairy and suckler heifers (Boissy & Le Neindre, 1997) and in Limousin x Jersey cross heifers and steers (Fisher *et al.*, 2000). Repeatability of the behavioural responses was examined in two of the studies. Hopster and Blokhuis (1993) repeated the social isolation procedure within a week, and found high repeatability values of between 0.58 and 0.69 for several behavioural measures. Fisher *et al.* (2000) repeated a test of sociality three times on the same animals at monthly intervals, and found the measure to be only moderately repeatable (0.34 ± 0.04).

The findings reported above indicate that variation between individuals should be seen in all of these tests. Repeatability has been demonstrated for three of the four tests. However, these findings come from animals of various breeds and ages that have been reared in different types of systems. No studies have examined these criteria on results from a number of tests carried out on the same animals. Hence, the objectives of this experiment were to carry out the four behaviour tests on a group of bull calves and a group of heifer calves, and examine the inter-animal variation and intra-animal repeatability of the data obtained.

3.2. Methods

3.2.1. The Tests

Four behavioural tests, the FF Test, the SS Test, the NO Test and the HA Test, were carried out on a group of 49 bull calves and a group of 56 heifers born in Year 1 (see Table 2.7). Each test was carried out three times on each animal, except the NO Test, which was only carried out once. To enable repeatability to be calculated for the NO Test also, it was carried out twice on the 72 heifers born in Year 2. Hence, results from all the tests carried out on the Year 1 bulls and heifers, and results from the NO Tests carried out on the Year 2 heifers, are reported here.

The test procedures were carried out as described in Chapter 2.3, and are only briefly described here. All the tests were carried out in the animals' home pens. The FF Test was a flight test carried out while the animal was standing at a feeder. The observer walked slowly towards the animal and its response was scored (see Chapter 2.3.2). The SS Test measured the animals' reactions to being put in the pen on their own for five minutes (Chapter 2.3.3). The NO Test measured the animals' responses to being put in the pen with a novel object for five minutes (Chapter 2.3.4). A traffic cone was used in the NO Tests carried out on the Year 1 animals, and for the first NO Test carried out on the Year 2 heifers. A plastic barrel was used in the second NO Test on the Year 2 heifers. Finally, the HA Test measured the time taken to move each animal in turn into a corner of the pen and hold it there for a 30-second period (see Chapter 2.3.5).

3.2.2. The Test Schedules

3.2.2.1. Year 1 Bulls

The test schedule carried out on the bull calves was described fully in Chapter 2.4.2.1 and Table 2.8, and only a brief reminder is noted here. The HA Test was carried out first, when the bulls were approximately eight months old. It was repeated three times at three-day intervals, with two non-test days in between. One and a half months later, one NO Test and three SS Tests were carried out on consecutive days. Finally FF Tests were carried out approximately one month later again, when each animal was tested two or three times over the course of nine days.

3.2.2.2. Year 1 Heifers

The test schedule carried out on the heifer calves was described fully in Chapter 2.4.2.2 and summarised in Table 2.9. Briefly, FF Tests were carried out first, when the heifers were approximately 11 months of age. The other tests were carried out just over a month later. A regime of three SS Test days, one non-test day, one NO Test day, three non-test days and finally three HA Test days was followed.

3.2.2.3. Year 2 Heifers

The test schedule carried out on the heifer calves born in Year 2 was similar to that followed for the Year 1 animals, and began when the heifers were approximately 12 months of age. It is described fully in Chapter 2.4.3.2 and Table 2.12. Only the NO Tests from the schedule are of interest for this chapter. Two NO Tests were carried out on consecutive days, after FF and SS Tests, and before HA Tests.

3.2.3. Behavioural Measures

The behavioural measures used to assess the animals' reactions to each of the tests are described in Chapter 2.3, and are briefly listed here. After carrying out each FF Test, the animal was given a score of 1 - 6 from the ordinal scale shown in Table 2.2. The score given depended on how quickly the calf moved away from the observer.

The SS, NO and HA Tests were recorded on videotape, which allowed total durations, frequencies and latencies of behaviours shown by the animal to be calculated. The total durations of nine behavioural states and the frequency of three events (listed in Table 2.3) were measured from the SS Tests. The main measures taken in the NO Test were 'Latency to touch the novel object' and 'Time spent in contact with the novel object' (Chapter 2.3.4). The main measure used in the HA Test (Chapter 2.3.5) was 'Latency to stand in the test corner for 30 seconds'. Two other variables were also examined, 'Percentage of test time spent standing in the test corner', and 'Percentage of test time spent running or walking'. A second method of analysis was also carried out on data from the HA Tests, as described in Chapter 2.3.5. This method involved calculating a 'Docility Score' from a range of variables, for example, aggressiveness displayed during different parts of the test, and total time spent in the corner (Le Neindre *et al.*, 1995). The two methods of analysis were compared to see which gave the most useful data in terms of intra-animal repeatability and inter-animal variation.

3.2.4. Statistical Analysis

The distribution and range of each test measure was examined for inter-animal variation in response. Principal Components Analysis (PCA) was used to look for the most appropriate behaviours by which to score the animals for their reactions in the SS Test. In most cases the data were not normally distributed and could not be transformed. Differences between the mean scores of all the animals obtained for each repeat of the test were examined using Paired t-tests. Restricted Maximum Likelihood (REML) analysis was carried out to estimate the variance components and deviances associated with the animals and test repeats. Repeatability values were calculated from a ratio of the variance components, and deviance differences were calculated to estimate the statistical significance of the values, as described in Chapter 2.5.2. Spearman Rank correlation coefficients were also calculated to investigate the relationship between scores from repeats of the tests, and were useful to double-check the repeatability coefficients from data that was not normally distributed.

3.3. Results

3.3.1. Flight-from-Feeder Test

3.3.1.1. Bulls

A wide range of response to the approaching observer was shown by the bull calves (Fig. 3.1). The flight scores obtained were normally distributed and covered the range 1 - 6. One animal was not observed to approach the feeders during the test periods and so was not tested. The remaining 48 animals were tested twice and 41 of these were also tested a third time.

The mean scores obtained by all the animals in each repeat of the test were compared, to see if there was an overall change in response to the test. There was no overall change in response between Tests 1 and 2, or between Tests 2 and 3 (Paired t-test: Tests 1 & 2, $n = 48$, mean difference = 0.25, $t = 1.11$, $p = 0.27$; Tests 2 & 3, $n = 41$, mean difference = 0.32, $t = 1.36$, $p = 0.18$).

Relationships between the animals' scores in repeated tests were examined by calculating correlation coefficients between pairs of tests. Moderate correlations between Tests 1 and 2, and between Tests 2 and 3 were found, indicating substantial relationships (Spearman Rank correlations; Tests 1 & 2, $n = 48$, $r_s = 0.46$, $p = 0.001$; Tests 2 & 3, $n = 41$, $r_s = 0.53$, $p < 0.001$).

A linear mixed REML model with 'animal number' and 'test repeat' fitted as random effects was used to obtain variance components, from which the repeatability was calculated. The model was run using different numbers of test repeats. Deviance differences were also calculated. Across the three test repeats the repeatability of the FF score was 0.52 ± 0.09 ($n = 41$, deviance difference = 30.98, $p < 0.001$). When calculated over just the first two tests, the repeatability coefficient obtained was slightly lower than over three tests (0.47 ± 0.11; $n = 48$, deviance difference = 11.67, $p < 0.001$). When calculated over the second two tests, it was slightly higher (0.53 ± 0.11; $n = 41$, deviance difference 14.84, $p < 0.001$). The correlation coefficients and repeatability estimates both indicate that the greatest reliability was seen between Tests 2 and 3.

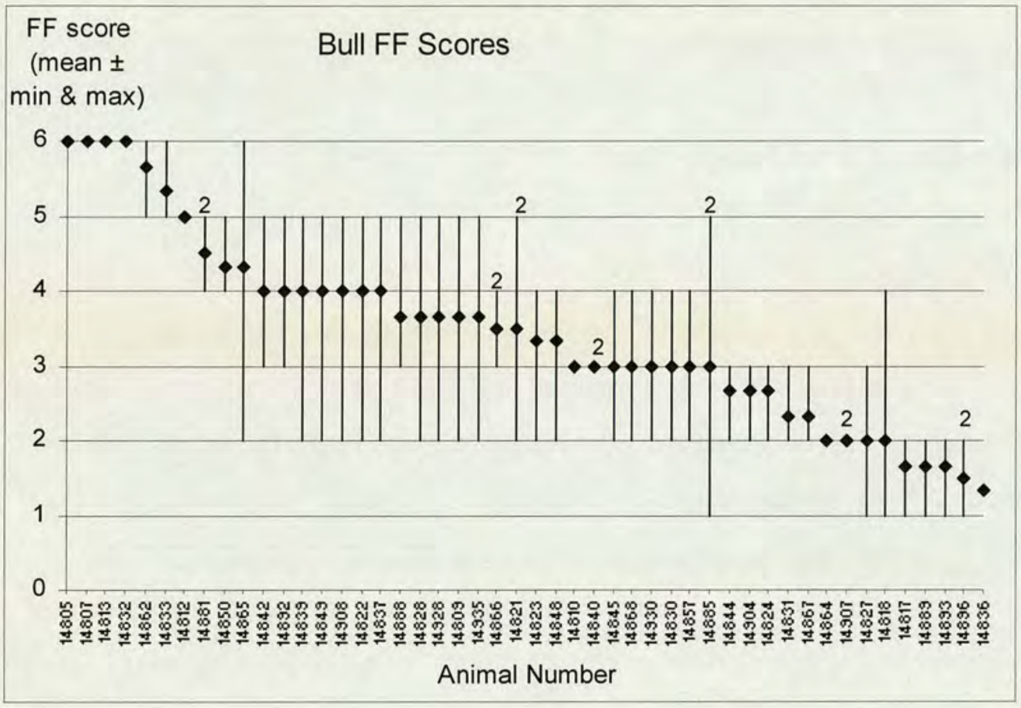


Figure 3.1: Flight-from-Feeder scores for 48 Year 1 bulls, in order of mean score. The means are for three repeats of the test, except where indicated by a '2'. The categories range from 1 (animal moved away when observer was within 2.00 - 1.25m) to 6 (animal didn't move back when touched). See Table 2.2 for the full scale.

3.3.1.2. Heifers

The mean flight scores obtained by the heifers, along with their minimum and maximum scores, are shown in Fig. 3.2. Two heifers were not seen at the feeders and one was tested only once and excluded from the analysis, giving a total of 53 heifers. All of these were tested twice, and 49 of them were tested a third time. A wide range of responses was seen, and the scores ranged from 1 - 6. As 17 of the 53 calves had a mean value of 6, the maximum score, the distribution is negatively skewed.

The mean scores obtained by the animals were compared between test repeats. There was a significant decrease in the mean score between Tests 1 and 2, but not between Tests 2 and 3 (Paired t-test: Tests 1 & 2, $n = 53$, mean difference = 0.491, $t = 3.40$, $p = 0.001$; Tests 2 & 3, $n = 49$, mean difference = 0.224, $t = 1.13$, $p = 0.263$). After being approached once, the heifers moved away more quickly in subsequent tests.

Despite the decrease in score between Tests 1 and 2, there was a strong correlation between the two tests (Spearman Rank correlations; $n = 53$, $r_s = 0.79$, $p < 0.001$). The correlation between scores in Tests 2 and 3 was modest ($n = 49$, $r_s = 0.64$, $p < 0.001$). These correlation values were likely to have been heavily influenced by the large number of animals scoring 6 on repeated tests.

Together, these results show there was a high correlation between the scores from Test 1 and Test 2, with the calves becoming consistently more fearful of human approach in Test 2. Although there was no difference in mean score between Tests 2 and 3, the correlation between these two repeats was weaker.

REML calculations were carried out on the data to obtain deviance difference and repeatability coefficients. A mixed model with 'animal number' and 'test repeat' fitted as random effects was used. Over the three tests, the repeatability was 0.50 ± 0.08 ($n = 49$, deviance difference = 33.53, $p < 0.001$). As expected from the correlation values, the highest repeatability of 0.58 ± 0.09 ($n = 53$, deviance difference = 21.53, $p < 0.001$) was found between Tests 1 and 2. The repeatability between Tests 2 and 3 was also high at 0.52 ± 0.11 ($n = 49$, deviance difference = 15.12, $p < 0.001$). The values obtained for the correlation coefficients and repeatability estimates back each other up well.

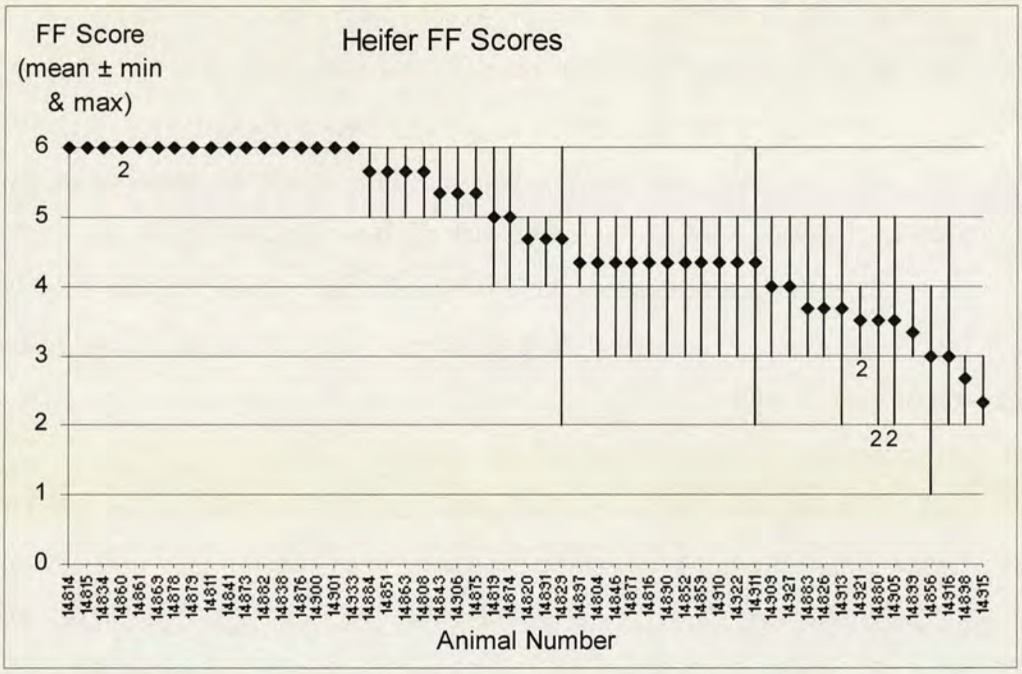


Figure 3.2: Flight-from-Feeder scores for 53 Year 1 heifers, in order of mean score. The means are for three repeats of the test, except where indicated by a '2'. The categories range from 1 (animal moved away when observer was within 2.00 - 1.25m) to 6 (animal didn't move back when touched). See Table 2.2 for the full scale.

3.3.2. Social Separation Test

3.3.2.1. Preliminary analysis - Bulls and Heifers

The behaviours seen in the SS Test were recorded as detailed in the ethogram in Table 2.3. The total durations of eight behaviour states (Escape (E), Gambol (G), Kneel (K), Run (R), Stand alert (SA), Stand occupied (S), Stand and sniff/lick/rub boards (B), Walk (W)), and the frequency of 'Vocalise' (V) were then analysed using Principal Components Analysis (PCA), in order to produce a graphical presentation of clusters of behaviours that may share a common motivational background (see 2.5.4). The state of Lying and the frequency of Urination and Defecation occurred with too low a frequency to be included in the analysis. Once useful groups of behaviours were identified, the inter-individual variation and intra-individual repeatability of these groups were examined.

Eleven of the bulls did not have data from all three SS Tests. Some of these animals were used in trial tests until the procedure was finalised on the first day. Some were lame and could not be tested, and two tried to jump out of the pen so persistently that their tests were stopped early. PCA was carried out on the data from the 38 bulls that were tested in all three test repeats. The eight durations and one frequency listed above from each test were used, giving 27 variables. The correlation matrix of the variables was used to calculate the eigenvectors (loadings) of the variables on each principal component. A plot of the loadings of the 1st two components, which described 25.0 % and 11.3 % of the total variation respectively, is shown in Fig. 3.3a. The three values for each behaviour, representing loadings for each of the three test repeats, have been circled, except R and G, which are quite spread out.

The plot gives an idea of whether the behavioural measures taken in the test are repeatable, as if they are, the three loadings of each behaviour should clump together. The values of the test repeats of most of the behaviours, W, SA, S, B, K and V clump closely together, with the repeats of R, E and G lying more spread out. The values of Tests 2 and 3 for E and R lie closer together than with Test 1, suggesting that a higher repeatability may be seen between these two tests.

The plot shows which behaviours may be the most useful as regards understanding the underlying motivations, as the value of the loading represents the degree to which the behaviour influences the component. It also shows which behavioural measures are highly correlated as these variables clump together. Four behaviours, W, R, E and V clumped together on a strong negative loading on Component 1 (except R1) and a fairly strong positive loading on Component 2 (except E1). It was assumed that animals that show 'Escape' behaviours in the pen have a strong motivation to return to their penmates, and hence these other three behaviours that clump together with E are also likely to be indicative of a strong sociality motivation. K, G and B also lie together, showing a positive loading on both components (with the exception of G3). These behaviours may represent 'excitability' or 'activity'.

Six of the heifers had missing data for at least one of the three repeats of the test, either because they were demonstrating behavioural signs of oestrus at the time of testing and their test was therefore discounted, or because tests were disturbed. They were removed for this analysis. Data from the 50 heifers that were tested in all three test repeats were analysed in the same way as the bulls. A plot of the first two components explained 26.4 % and 11.2 % of the total variance respectively, very similar figures to the bull data. Again, behaviour loadings from the three test repeats lie closely together in one or both dimensions, indicating that the animals are showing consistent reactions when the test is repeated (see Fig. 3.3b).

Similarly to the bulls, W, E and R are clumped together in both dimensions. Again they show a strong negative loading on Component 1, and a fairly strong positive loading on Component 2 (except R3). In Component 1, V also clumps with them. Again, these behaviours seem likely to be indicative of a strong sociality motivation. B, K and G also cluster together, although they load negatively on Component 2 compared to positively in the bull data. S loads on to a fairly similar area to the bull data in Component 1, but loads strongly positively on Component 2 compared to strongly negatively in the bulls.

The PCA suggested that behaviours W, E and R were highly correlated, and strongly influence Component 1 in both the bulls and the heifers. The total durations of these three behaviours were taken to be a measure of sociality from this test, and further analysis was carried out on this combined measure 'WER'.

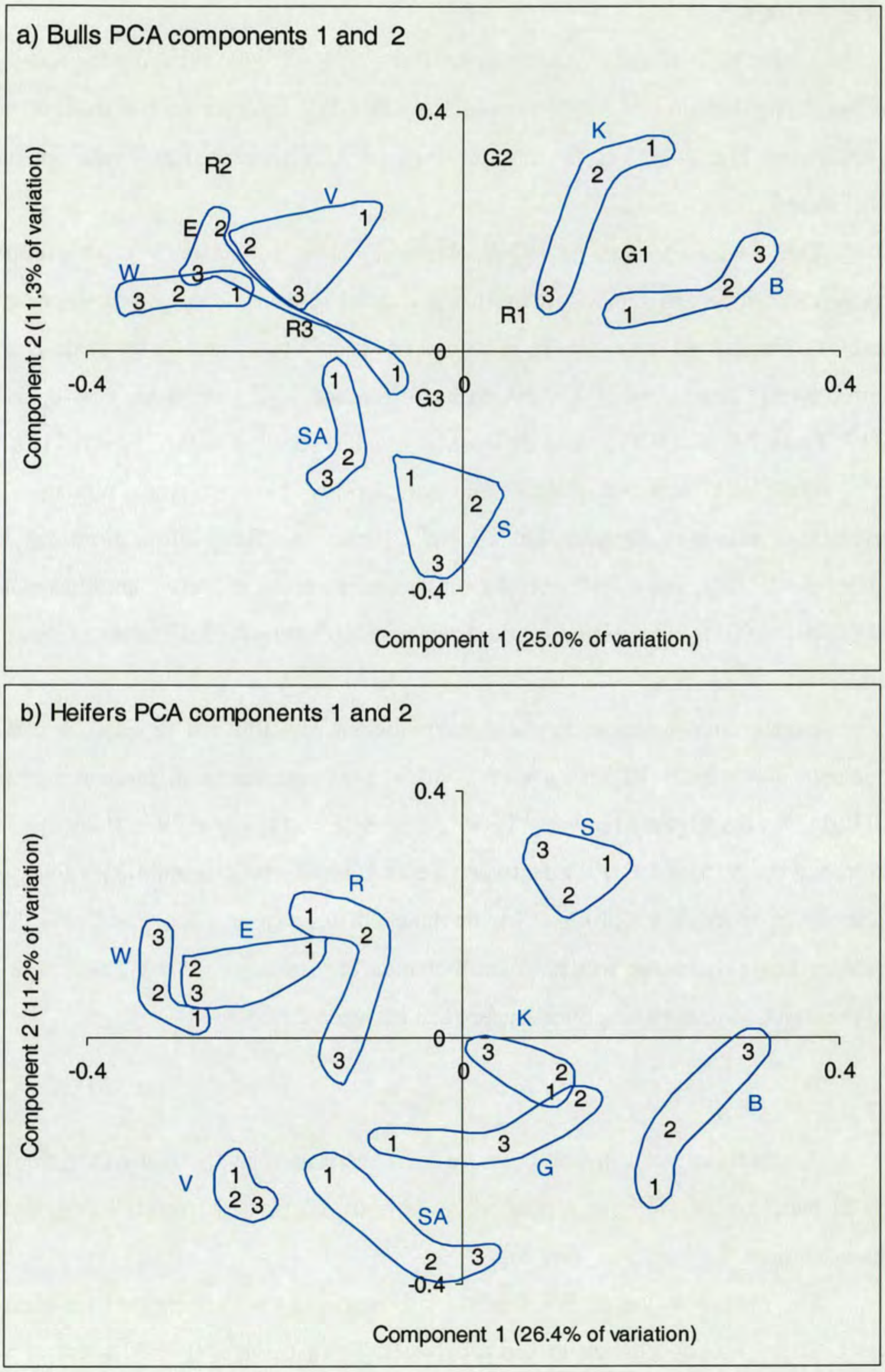


Figure 3.3: A Principal Components Analysis of data from the three SS Tests carried out on 38 bulls (a) and 50 heifers (b). The behaviours measured are durations of B = stand and sniff/lick/rub boards, E = escape, G = gambol, K = kneel, R = run, S = stand occupied, AS = stand alert, W = walk, and a frequency V = vocalise.

3.3.2.2. Bulls

All the bulls tested in one or more of the tests were included for the rest of the analysis. The duration of WER ranged from 0 - 217 seconds with a median of 31 seconds (see Fig. 3.4). The distribution was positively skewed and was not easily transformed.

The median values of WER obtained across all animals were compared between the three test repeats to see if the overall response changed as the test was repeated. No change was seen in response between Test 1 and 2, or Tests 2 and 3 (Paired t-test; Tests 1 & 2, $n = 39$, mean difference = -7.1 seconds, $t = -0.82$, $p = 0.417$; Tests 2 & 3, $n = 41$, mean difference = -2.0 seconds, $t = -0.37$, $p = 0.712$).

The data was examined for correlations between test repeats. Low correlations were seen between Tests 1 and 2 (Spearman Rank correlation; $n = 39$, $r_s = 0.36$, $p = 0.024$), and a high correlation was seen between Tests 2 and 3 ($n = 41$, $r_s = 0.81$, $p < 0.001$), indicating a marked relationship between WER scores from these tests.

Intra-animal repeatability was calculated across the test repeats. Across the three tests it was high (REML; $n = 43$, $r = 0.54 \pm 0.09$, deviance difference = 32.04, $p < 0.001$). It was highest between Tests 2 and 3 ($n = 41$, $r = 0.79 \pm 0.06$, deviance difference = 42.15, $p < 0.001$). Between Tests 1 and 2, the repeatability values were moderate ($n = 39$, $r = 0.40 \pm 0.14$, deviance difference = 6.55, $p < 0.05$). These results indicate that there was no overall change in response as the test was repeated, and the most consistent responses were seen between 2 and 3.

3.3.2.3. Heifers

All heifers tested in one or more tests were used in the rest of the analysis. The distribution of WER was positively skewed and ranged from 0 - 264 seconds with a median of 49 seconds (see Fig. 3.5).

The median value of WER across all the heifers was compared for each test repeat. There was no change in score between Tests 1 and 2 (Paired t-test; $n = 52$, mean difference = 3.2 seconds, $t = 0.40$, $p = 0.693$). There was an increase in score between Tests 2 and 3 ($n = 52$, mean difference = -15.9 seconds, $t = -2.82$, $p = 0.007$), implying higher levels of sociality were seen in the third test.

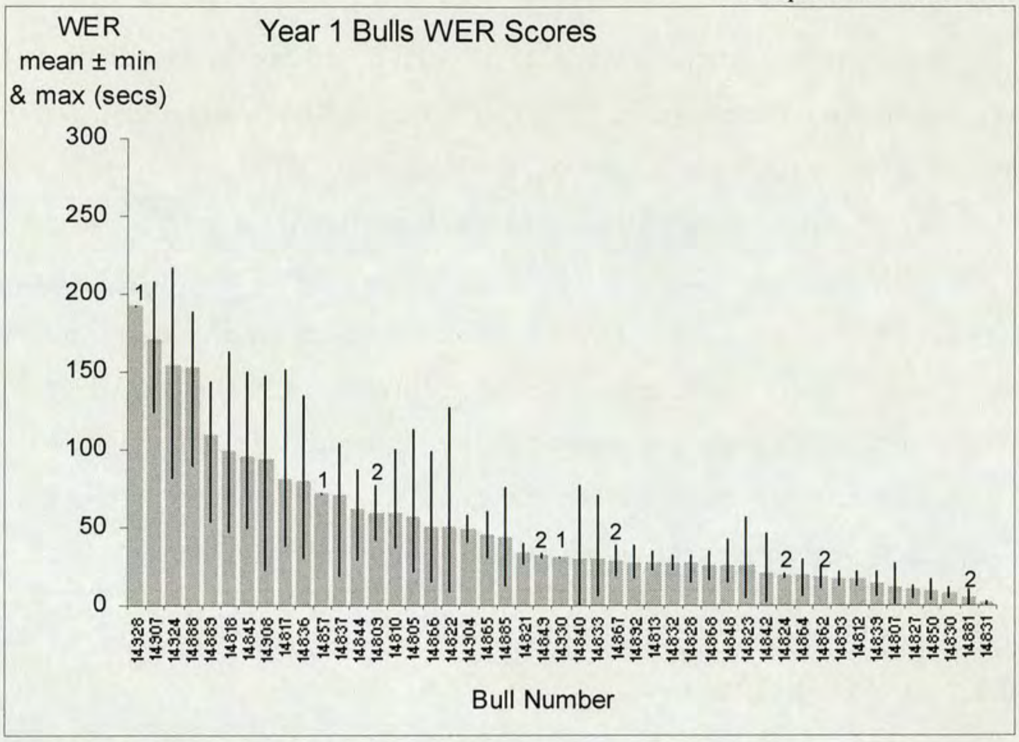


Figure 3.4: WER scores from the Social Separation Tests carried out on 47 Year 1 bulls. WER = the duration of test time spent Walking, attempting to Escape, or Running. The means are from three tests unless indicated by '2' or '1'.

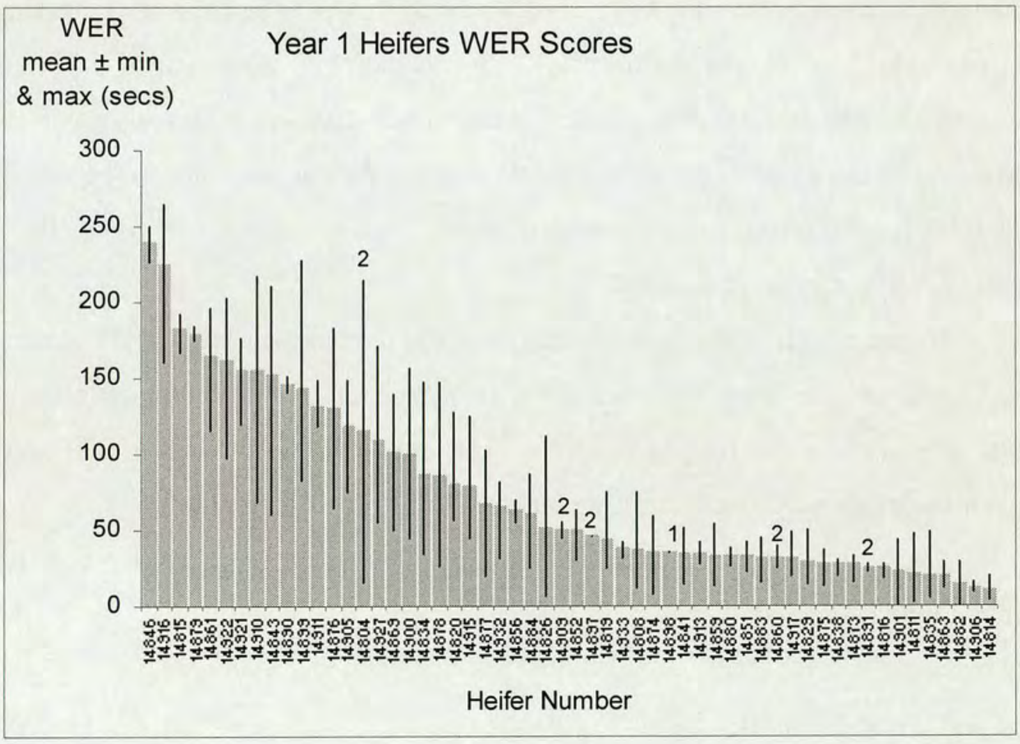


Figure 3.5: WER scores from the Social Separation Tests carried out on 56 Year 1 heifers. WER = the duration of test time spent Walking, attempting to Escape, or Running. The means are from three tests unless indicated by '2' or '1'.

A moderate relationship was seen between the scores obtained in Tests 1 and 2 (Spearman Rank correlation; $n = 52$, $r_s = 0.53$, $p < 0.001$), and a high correlation was seen between Tests 2 and 3 ($n = 52$, $r_s = 0.70$, $p < 0.001$).

Repeatability across all three tests was high (REML; $n = 50$, $r = 0.68 \pm 0.06$, deviance difference = 74.22, $p < 0.001$) and across pairs of tests was high or very high (Tests 1 & 2; $n = 52$, $r = 0.59 \pm 0.09$, deviance difference = 22.91, $p < 0.001$; Tests 2 & 3; $n = 52$, $r = 0.81 \pm 0.05$, deviance difference = 57.46, $p < 0.001$). Again, the most marked relationship was seen between scores from Tests 2 and 3.

These results indicate that the test scores became higher in the third repeat of the test, and a very strong relationship was seen between Tests 2 and 3.

3.3.3. Novel Object Test

3.3.3.1. Year 1 Bulls

Of the 49 bulls, two were lame and were not tested. The main measure taken, 'Latency', referred to the time taken for the animal to touch the cone after entering the test pen. The Latency ranged from 2 - 300 seconds with a median of 9.2 seconds. Most of the animals approached and touched the cone very quickly, and so the variation in Latency seen between most of the animals was very small (see Fig. 3.6a), and the distribution showed a positive skew. Two animals didn't touch the cone during the five-minute test period.

Because so little variation in response to the cone was seen, another measure was examined. The distribution of 'Time spent in contact with the novel object' was negatively skewed and ranged from 0 - 297 seconds with a median of 243 seconds. This measure showed wide variation between the animals (see Fig. 3.6b).

The test was only carried out once and therefore repeatability coefficients could not be estimated.

3.3.3.2. Year 1 Heifers

Six of the 56 heifers were not tested as they were demonstrating behavioural signs of oestrus or were sick on the day of testing. The Latency showed a range of 4 - 161 seconds and a median of 10 seconds, and had a highly positively skewed distribution, as most of the heifers approached the cone very quickly. As was the case with the bulls, there was not much variation in response seen between most of the animals (see Fig. 3.7a).

The distribution of 'Time spent in contact with the novel object' was negatively skewed and ranged from 38 - 287 seconds, with a median of 222 seconds. This measure showed wide variation between animals (see Fig. 3.7b).

The test was carried out once and therefore repeatability coefficients could not be estimated.

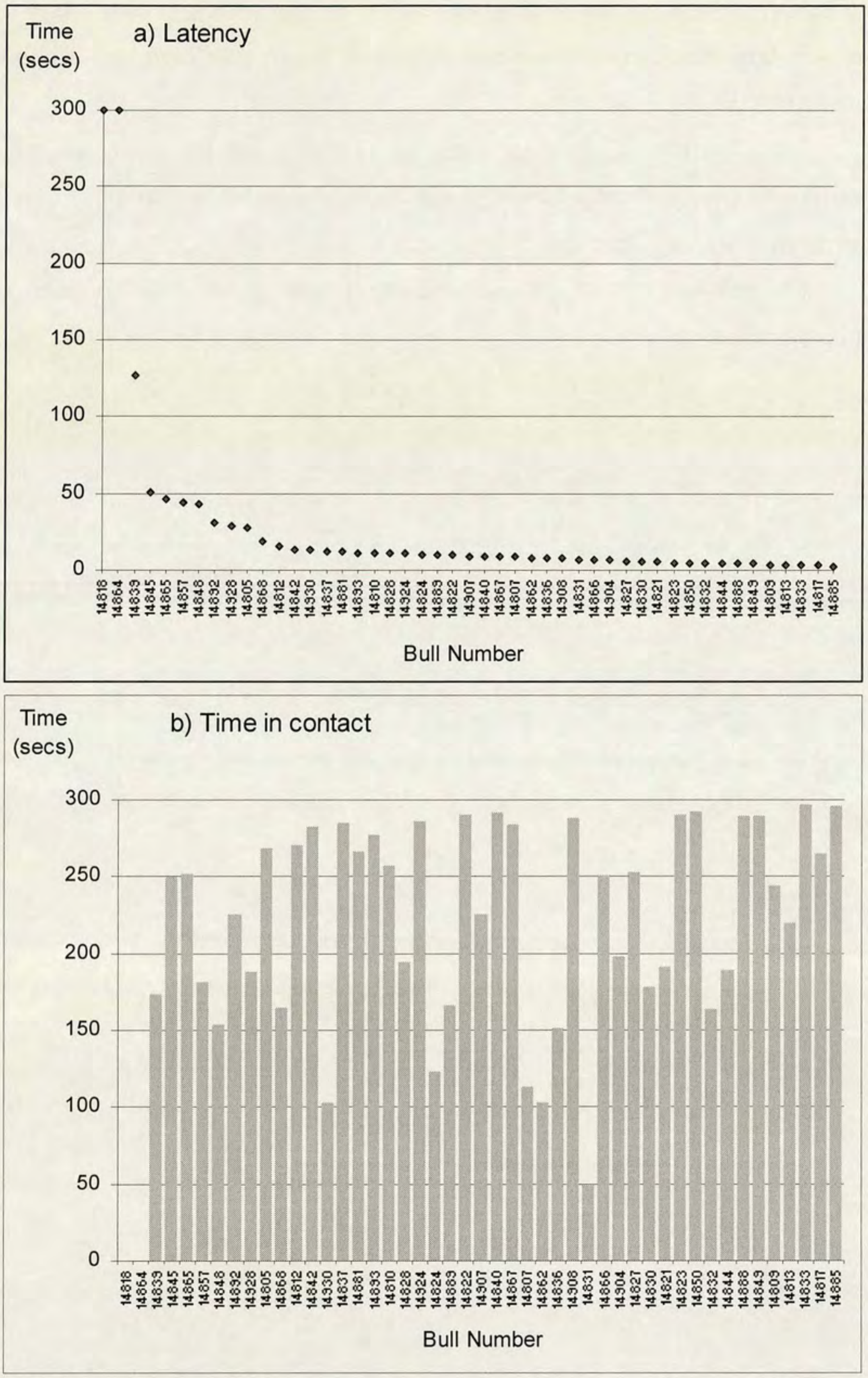


Fig. 3.6: Measures taken in the Novel Object Test carried out on the Year 1 bulls; a) Latency to touch the novel object, and b) Time spent in contact with novel object. The 47 animals are listed in order of decreasing latency.

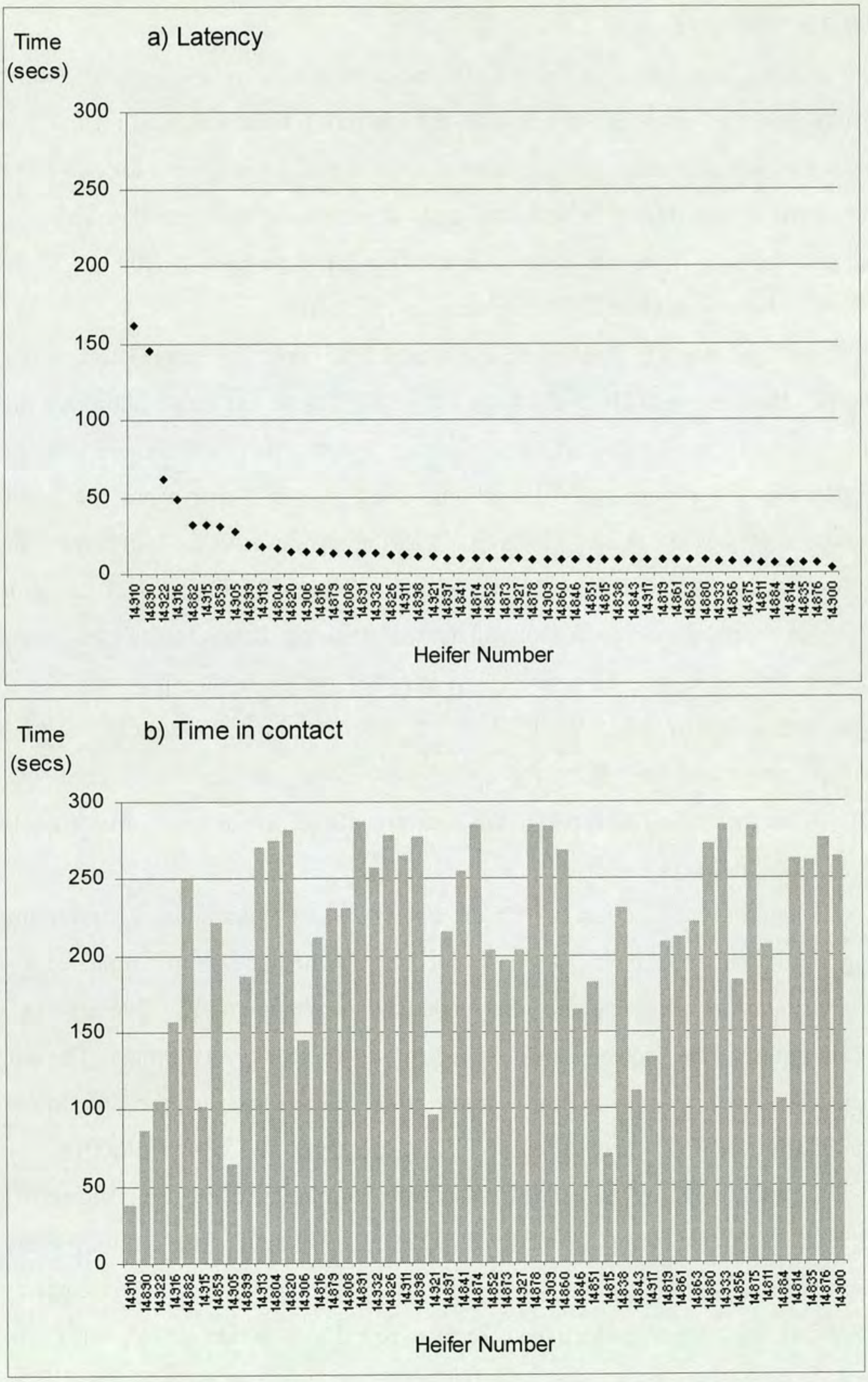


Fig. 3.7: Measures taken in the Novel Object Test carried out on the Year 1 heifers; a) Latency to touch the novel object, and b) Time spent in contact with novel object. The 50 animals are listed in order of decreasing latency.

3.3.3.3. Year 2 Heifers

The Year 2 heifers were tested twice, allowing relationships between the results from the two tests to be investigated. A traffic cone was used for the first test and a plastic barrel was used for the second. Data was not available for two heifers as they were demonstrating behavioural signs of oestrus on both test days. Nine others had data missing from one of the two tests for the same reason. This left 70 heifers with data for one test and 61 heifers with data for both.

All the animals approached and touched the cone. The pattern seen in the data was similar to that shown by the Year 1 animals. Latency in Test 1 showed a median of 4.5 seconds, and a range of 1 - 96 seconds. Latency in Test 2 showed a median of 3.0 seconds and a range of 2 - 10 seconds (see Fig. 3.8a). Apart from one heifer that showed a very long Latency in the first test, the responses to the barrel were similar to those to the cone, and very little variation in response to either object was seen.

A correlation coefficient calculated between Latencies in the two tests showed that there was no relationship between the scores in the tests (Spearman Rank correlation; $n = 61$, $r_s = 0.04$, $p = 0.775$), and, correspondingly, the measure was not repeatable (REML; $n = 61$, $r = -0.04 \pm 0.13$, deviance difference = 0, $p > 0.05$). This lack of a relationship was symptomatic of the lack of variation between the animals in Latency.

'Time spent in contact with the object' did show variation between animals (Fig. 3.8b). The median score was 97 seconds for the cone, with a range of 4 - 291 seconds, and for the barrel was 198 seconds, with a range of 12 - 296 seconds. The distributions of the scores from neither test were normally distributed. The animals spent more time in contact with the barrel in the second test than the cone in the first (Paired t-test; $n = 61$, mean difference = -64.4 seconds, $t = -5.88$, $p < 0.001$).

The calculation of a Spearman Rank correlation coefficient showed that a moderate relationship was seen between the contact shown by the animals in the two tests ($n = 61$, $r_s = 0.47$, $p < 0.001$). The repeatability of 'Time spent in contact with the object' was also calculated, and was also found to be moderate ($n = 61$, $r = 0.26 \pm 0.12$, deviance difference = 4.40, $p < 0.05$).

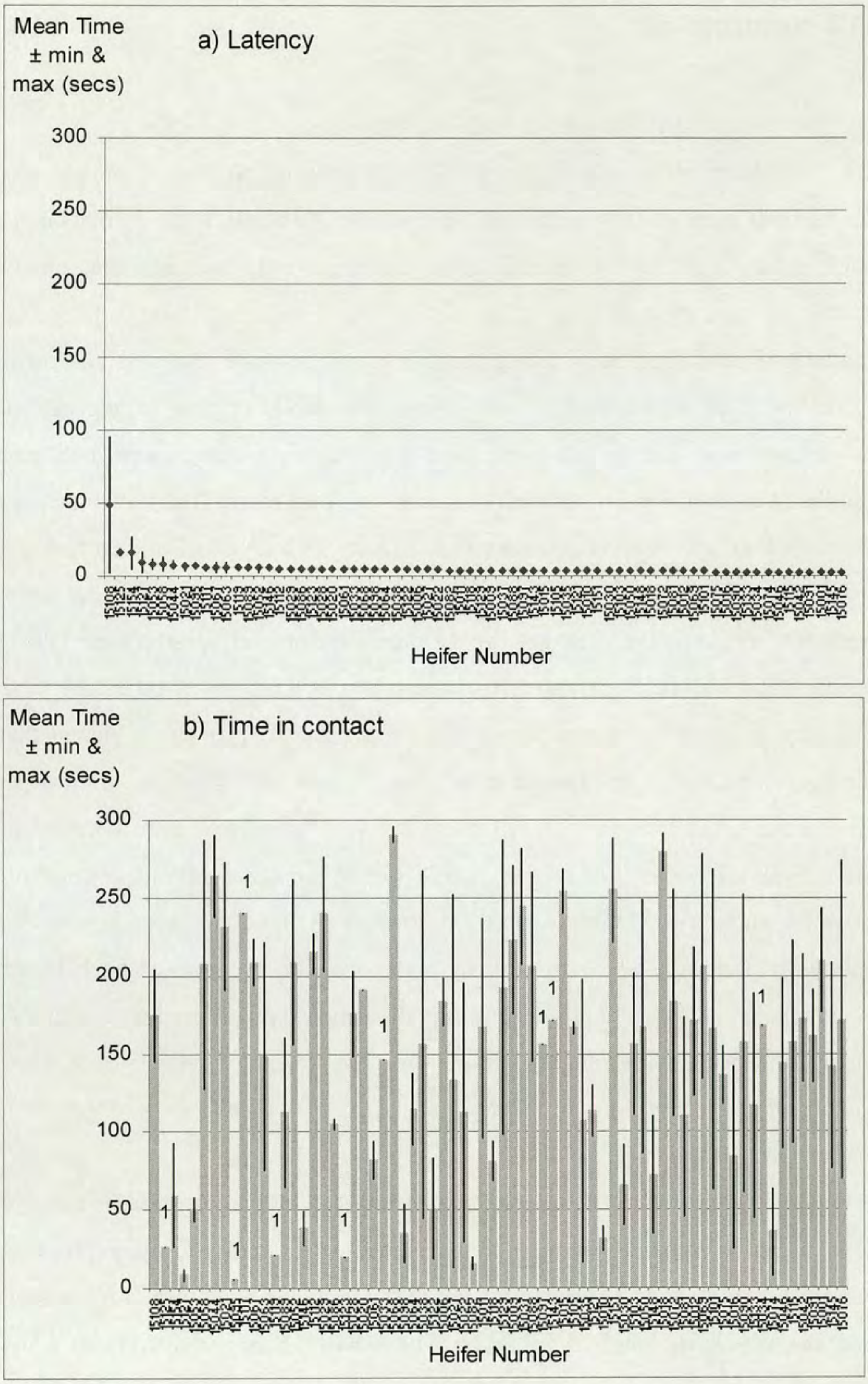


Fig. 3.8: Mean scores from two Novel Object Tests carried out on 70 Year 2 heifers; a) Latency to touch the object, and b) Time spent in contact with object. The heifers are listed by decreasing latency. Heifers tested once are marked with a '1' in graph b.

3.3.4. Handling Test

3.3.4.1. Preliminary Analysis and Discussion - Bulls and Heifers

As there were two possible methods of analysing the data from this test, the two methods were initially compared on approximately half of the animals in a pilot analysis. Using half of the animals allowed a decision to be reached as quickly as possible on how best to analyse the rest of the test videos. Firstly, latencies and durations of behaviours seen during part iii of the tests were analysed, and secondly, a 'Docility Score' was calculated using several variables relating to general activity and performance. The results from these two analysis methods were compared as described below, and the most useful analysis was then carried out on all the animals.

Bull groups B and C were analysed giving 28 animals. The data from two of the animals were missing for the third repeat of the test, as one of the tests was interrupted and the other was not carried out as the animal became lame. The main measure taken in the handling part of the test was 'Latency to be held in the test corner for 30 seconds' (hereafter referred to as 'Latency'). It was assumed that the animals that had not been retained in the corner by the end of the two-minute period, and therefore had latencies of 120 seconds, would not have been retained in the corner if the test had carried on for a longer period. To deal with this in a statistically sensible manner, 30 seconds was added on to the Latency of these animals for all analyses of the data. The aim of this was to increase the distinction between the animals that could be held in the corner by the handler, and those that could not. The animals showed a wide range of responses. The Latencies ranged from 12 - 150 seconds and were not normally distributed, as 21 of the animals were not cornered during at least one test.

The Docility Scores were calculated for each animal as described in Chapter 2.3.5, using a computer programme designed specifically for this purpose (Le Neindre *et al.*, 1995). This method involves calculating a score from a range of variables, which are listed in Table 2.6. The Docility Scores ranged from 7.5 - 17.0 (from possible scores of 6.5 - 17.0) and also were not normally distributed. A comparison of the means of the two measures for each calf across the three tests (Fig. 3.9) shows that their distributions were very similar. The mean Latency for each animal shows a wider range from within the possible range of scores than the mean Docility Scores. The two measures are very highly negatively correlated (Spearman Rank correlation: $n = 28$, $r_s = -0.96$, $p < 0.001$), showing that animals with a high Docility Score were quicker to move into the corner and be kept there by the handler. This suggests that many of the behavioural variables used in the calculation of the Docility Score may be highly correlated with Latency.

The repeatability of each measure was calculated across the three repeats of the tests, and across pairs of repeats, and these values are summarised in Table 3.1. Over three test repeats, the repeatability of the Docility Score was moderate (REML; $r = 0.26 \pm 0.13$), but of the Latency was not significantly different from zero. There was little consistency in either measure between Tests 1 and 2. A consistent Docility Score was seen between Tests 2 and 3, with a moderate repeatability of 0.46 ± 0.15, but Latency was again not repeatable. Latency was not repeatable across any of the test repeats, and Docility was more repeatable between Tests 2 and 3 than between all the tests or Tests 1 and 2.

Table 3.1: Repeatability values and deviances differences of two measures taken in the HA test, on a preliminary sample of bulls and heifers.

		'Latency to stand in corner'		Docility Score	
Bulls	Number	Repeatability ± s.e.	Deviance Difference	Repeatability ± s.e.	Deviance Difference
Tests 1, 2 & 3	26	0.20 ± 0.12	3.14 ^{NS}	0.26 ± 0.13	4.98 *
Tests 1 & 2	28	0.003 ± 0.19	0.003 ^{NS}	0.01 ± 0.19	0 ^{NS}
Tests 2 & 3	26	0.32 ± 0.17	3.02 ^{NS}	0.46 ± 0.15	6.54**
Heifers					
Tests 1, 2 & 3	26	0.45 ± 0.12	14.57 ***	0.35 ± 0.13	8.54 **
Tests 1 & 2	27	0.44 ± 0.16	5.72 *	0.40 ± 0.16	4.70 *
Tests 2 & 3	26	0.57 ± 0.13	10.56 **	0.30 ± 0.18	2.41 ^{NS}

*** p < 0.001, ** p<0.01, * p<0.05, ^{NS} p>0.05.

This preliminary analysis was also carried out on approximately half of the heifers. The data from groups C and D were analysed giving 26 animals with three test repeats and one animal with two repeats. A wide range of response (7 - 150 seconds) was seen between the calves for the Latency, which was not normally distributed (Fig. 3.10a). A smaller range was seen for the Docility Score (9.25 - 16.9), which also did not have a normal distribution (Fig. 3.10b). A very high negative correlation between the two measures was seen (Spearman Rank correlation; $n = 27$, $r = -0.94$, $p < 0.001$), again implying that variables used in the calculation of the Docility Score must be correlated, and that Latency has a big influence on the calculation of the Docility Score.

The repeatability of the two measures across successive test repeats was calculated and are summarised in Table 3.1. Overall, the repeatability values were moderate to high, and higher for Latency than for the Docility Score. A greater degree of repeatability of Latency was seen between Tests 2 and 3 ($r = 0.57 \pm 0.13$) than between all three tests ($r = 0.45 \pm 0.12$), or between Tests 1 and 2 ($r = 0.44 \pm 0.16$). In contrast, the highest degree of repeatability of Docility Score was seen across Tests 1 and 2 ($r = 0.40 \pm 0.16$), and across Tests 2 and 3 was not significant.

Differences were seen between the results from the bulls and heifers. The Latency measure gave *lower* repeatability values than the Docility Scores for the bulls, but *higher* repeatability values than the Docility Scores for the heifers. Overall, repeatability values calculated were generally higher for the heifers than for the bulls. Two methodological differences may have contributed to the differences. Firstly, the HA Test was carried out *before* the other behaviour tests with the bulls, but *after* the other tests in the heifers. Secondly, a gap of two days was left in between each test for the bulls, but the heifer tests were carried out on successive days. These points are discussed further in 3.4.4.

For both the bull and heifer data, the very strong correlations between the two measures demonstrated that the calculation of the Docility Score is influenced highly by the latency measure. On the basis of these results, it was decided to use the Latency measure for the full analysis on the full number of animals, rather than using the Docility program, which combines several behavioural variables and assigns arbitrary weights to each. Durations of other behaviours seen during the test were also recorded and analysed, and details of these follow.

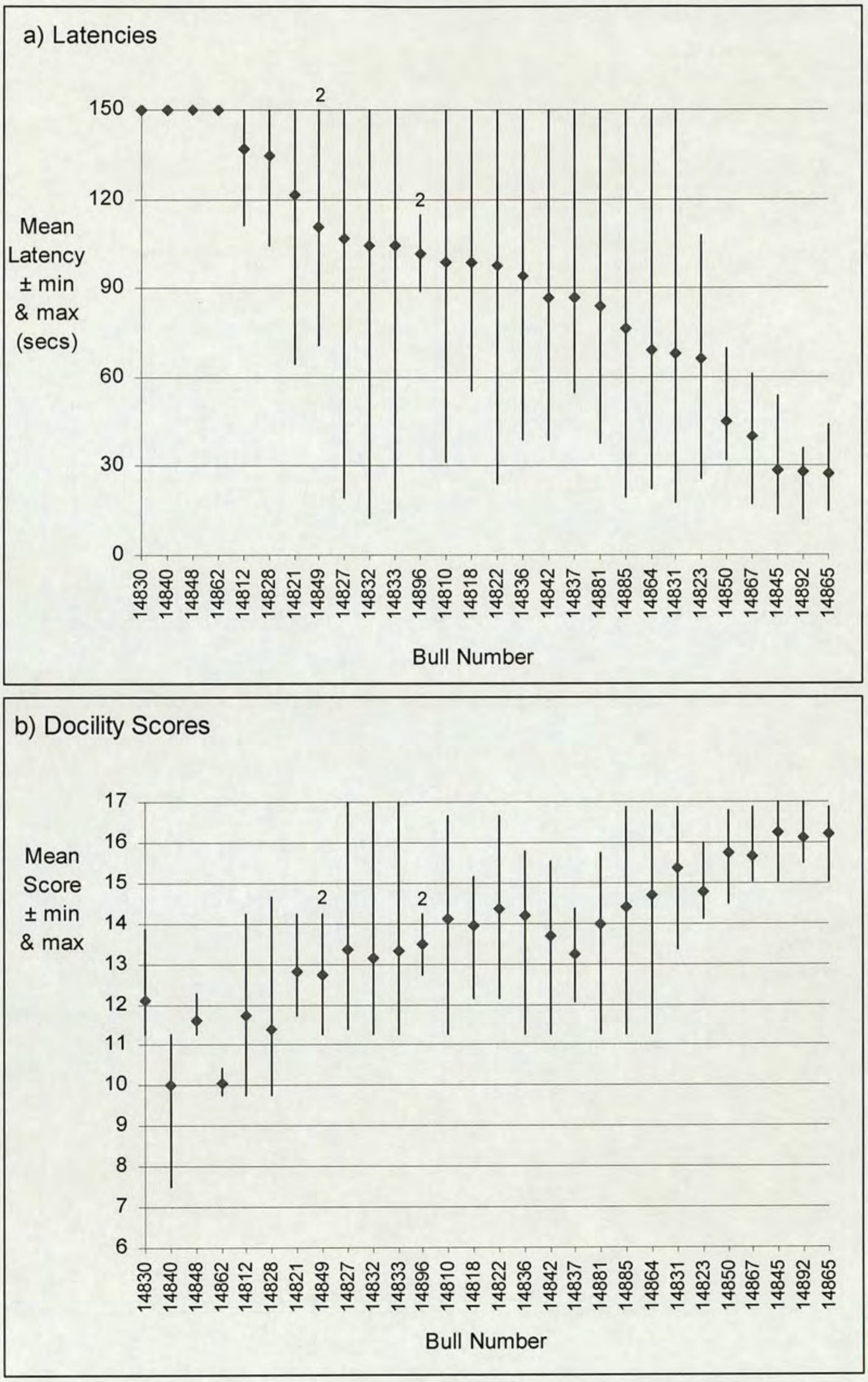


Fig. 3.9: Mean scores from three Handling Tests carried out on 28 Year 1 bulls; a) Latency to stand in the test corner for 30 seconds, and b) Docility Score, calculated from a range of variables. The animals are listed in order of decreasing Latency. The means are from three tests unless indicated by a '2'.

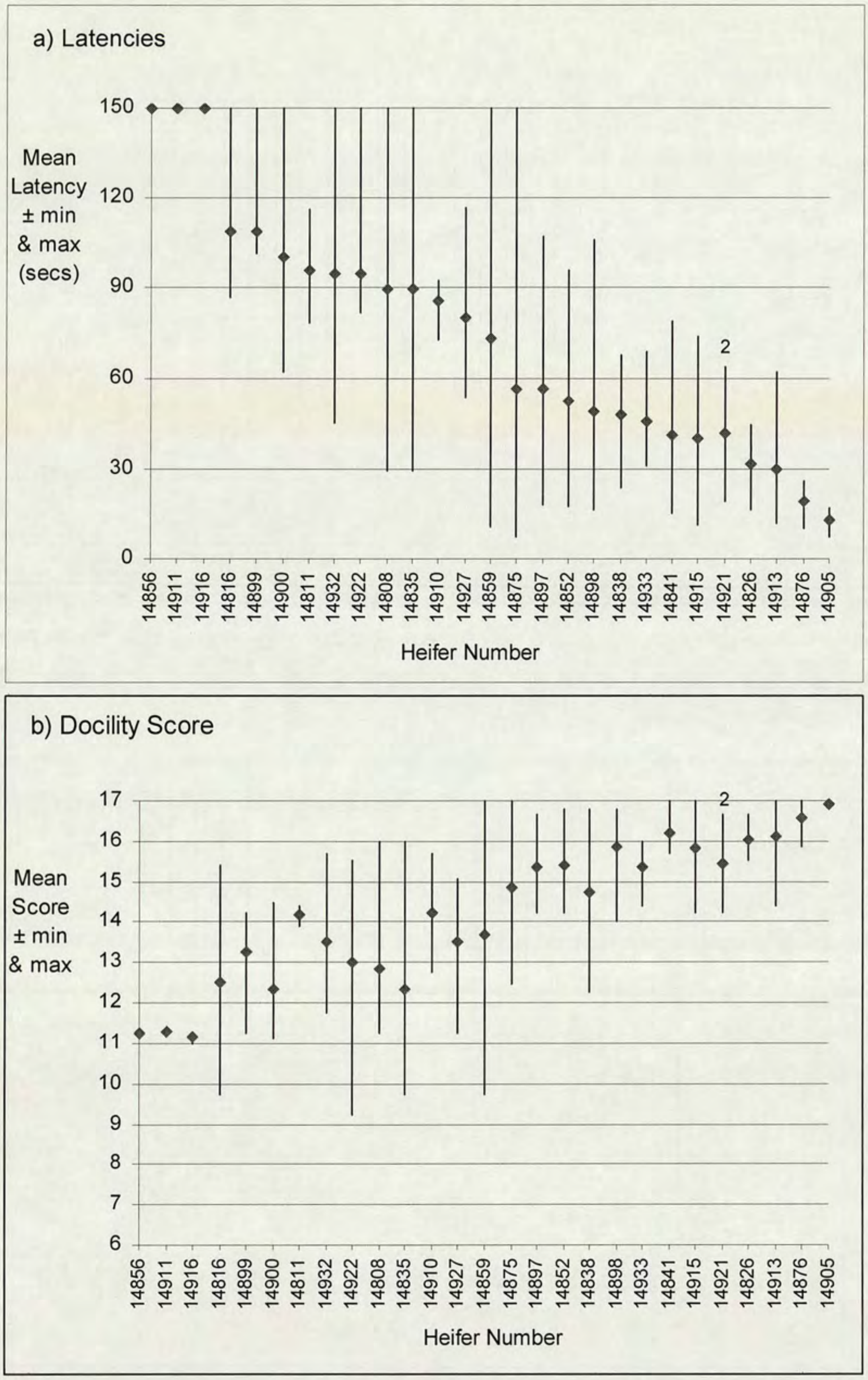


Fig. 3.10: Mean scores from three Handling Tests carried out on 27 Year 1 heifers; a) Latency to stand in the test corner for 30 seconds, and b) Docility Score, calculated from a range of variables. The animals are listed in order of decreasing Latency. The means are from three tests except the animal marked by a '2'.

3.3.4.2. Full Analysis - Bulls

The analysis described in 3.3.4.1 was repeated on data from all the bulls. Of the 49 animals, one was not tested at all, as it was missed on the first test day. Data from the first tests carried out on four animals were missing, as these were used as a pilot trial. One animal was not tested a third time as it was lame, and the test of another was stopped early because it was showing dangerously aggressive behaviour. Again, 30s were added onto the latency of animals that had not been kept in the corner by the end of the test period. The mean 'Latency to stand in the corner for 30 seconds' ranged from 28 to 150 seconds (Fig. 3.11). The Latencies were not normally distributed, as 33 of the animals had a Latency of 150 seconds during at least one test.

The mean Latency of all the animals obtained in the test repeats were compared. No differences were found between Tests 1 and 2, or Tests 2 and 3 (Paired t-test; Tests 1+2, $n = 44$, mean difference = -1.7 seconds, $t = -0.15$, $p = 0.881$; Tests 2+3, $n = 46$, mean difference = -11.13 seconds, $t = -1.17$, $p = 0.249$).

Consistency of response across test repeats by individual bulls was examined. No relationship was found between the Latencies seen in Test 1 and 2 (Spearman Rank correlation; $n = 44$, $r_s = 0.02$, $p = 0.897$). A low correlation was found between Tests 2 and 3 ($n = 46$, $r_s = 0.34$, $p = 0.021$). Repeatability coefficients were also calculated. No repeatability was seen over the three tests (REML; $n = 42$, $r = 0.14 \pm 0.10$; deviance difference = 2.42, $p > 0.1$) or between Tests 1 and 2 ($n = 44$, $r = 0.02 \pm 0.15$, deviance difference = 0.01, $p > 0.1$). A moderate repeatability was seen between Tests 2 and 3 ($n = 46$, $r = 0.33 \pm 0.13$; deviance difference = 5.21, $p < 0.025$). Overall, these results show that the test response was found to be consistent between only Tests 2 and 3, and that this relationship was definite but small.

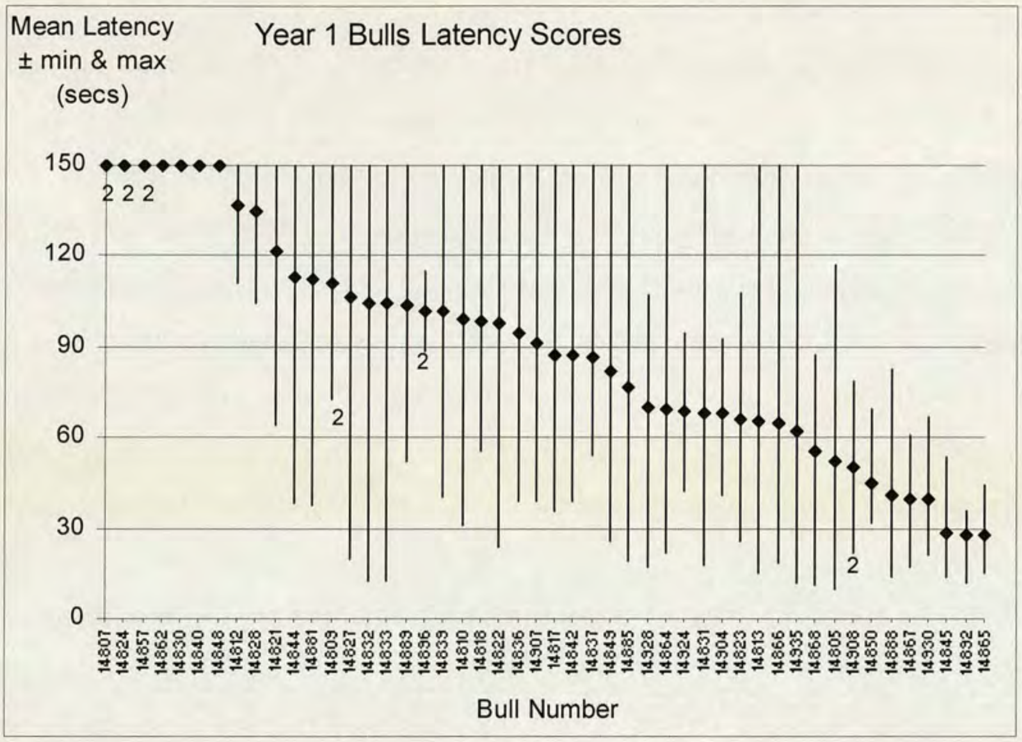


Fig. 3.11: Latency scores from the Handling Tests carried out on 48 Year 1 bulls. Latency = Time taken for the animal to move into the test corner for 30 seconds. The means are from three tests unless indicated by '2'.

'Percentage of test time spent standing in test corner' (hereafter called '% Corner Time') was examined. This variable ranged from 0 - 74.6 % and had a very skewed distribution with many zero values. Small, insignificant relationships were found between pairs of test repeats (Spearman Rank correlation; Tests 1 & 2, n = 44, $r_s = 0.12$, p = 0.443, Tests 2 & 3, n = 46, $r_s = 0.27$, p = 0.063), showing that this measure was not repeatable across tests. A very high negative correlation was seen between this measure and Latency (n = 44, $r_s = -0.90$, $p < 0.001$), implying that this measure is not adding any additional information.

The percentage of time the animals spent running or walking ('% Run Walk') was also calculated to determine if the animals' activity levels in the tests were repeatable. The data, which ranged from 3 - 89 %, were approximately normally distributed. Again, only small, insignificant relationships were seen between tests (Spearman Rank correlation; Tests 1 & 2, n = 44, $r_s = 0.18$, p = 0.250, Tests 2 & 3, n = 46, $r_s = 0.29$, p = 0.050), showing that this measure was also not repeatable across tests.

Calculation of durations of time spent standing, walking and running during the *first* stage of the test were also made. This was the period before the handling, when the animal was allowed to settle for 30s with the handler standing outside the pen (see Chapter 2.3.5.). This data is described in Chapter 4.2.2, where it is used to compare the levels of activity with those seen in the SS Test, in order to establish if separation from their herdmates contributed any part to the behavioural response shown by the animals in the HA Test. It is not described further here.

3.3.4.3. Full Analysis - Heifers

The analysis carried out on data from the bulls was repeated on the full number of 56 heifers. Data from the second test repeat of one of the heifers was not included as it was demonstrating behavioural signs of oestrus at the time of testing. The mean Latencies of each animal across three tests ranged from 13 - 150 seconds (see Fig. 3.12). Over half of the animals had a Latency of 150 seconds during at least one test repeat, causing a skewed distribution.

The latencies obtained by the heifers in the successive test repeats were compared to see if there were overall changes in response to the test when carried out a second and third time. The mean Latency was shorter in Test 2 than in Test 1 (Paired t-test; $n = 55$, mean difference = 24.5 seconds, $t = 3.59$, $p = 0.001$), but there was no difference between Test 2 and Test 3 ($n = 55$, mean difference = -1.61 seconds, $t = -0.23$, $p = 0.819$).

Consistency of response across test repeats by individual heifers was examined. Substantial relationships were found between Latencies shown in Tests 1 and 2, and Tests 2 and 3 (Spearman Rank Correlation coefficients; Tests 1 & 2, $n = 55$, $r_s=0.56$, $p<0.001$; Tests 2 & 3, $n = 55$, $r_s = 0.49$, $p<0.001$). Repeatability coefficients were also calculated between pairs of tests, and over all three test repeats. A moderate repeatability was seen over all tests ($r = 0.43 \pm 0.08$, $n = 55$, deviance difference = 26.80, $p<0.001$), and between Tests 1 and 2 ($r = 0.47 \pm 0.11$, $n = 55$, deviance difference = 13.81, $p<0.001$). Tests 2 and 3 were highly repeatable (0.53 ± 0.97, $n = 55$, deviance difference = 17.87, $p<0.001$).

'Percentage of test time spent standing in test corner' ('% Corner Time') was examined. This variable ranged from 0 - 81 % and showed a very skewed distribution with many zero values. A substantial relationship was found between Tests 1 and 2 (Spearman Rank correlation, $r_s = 0.59$, $n = 55$, $p < 0.001$), and Tests 2 and 3 ($r_s = 0.43$, $n = 55$, $p = 0.001$). A very dependable relationship was found between this measure and Latency ($n = 54$, $r_s = -0.95$, $p < 0.001$).

The percentage of time the animals spent running or walking ('% Run Walk') was also calculated. The data ranged from 8 - 89 % and was not normally distributed. Relationships between repeat test measures were similar to '% Corner Time', with a substantial relationship seen between Tests 1 and 2 ($r_s = 0.54$, $n = 55$, $p < 0.001$) and 2 and 3 ($r_s = 0.46$, $n = 55$, $p < 0.001$).

Calculations of durations of time spent standing, walking and running during the *first* stage of the test were also made. This was the period before the handling, when the animal was allowed to settle for 30s with the handler standing outside the pen. This data is described in Chapter 4.2.2.

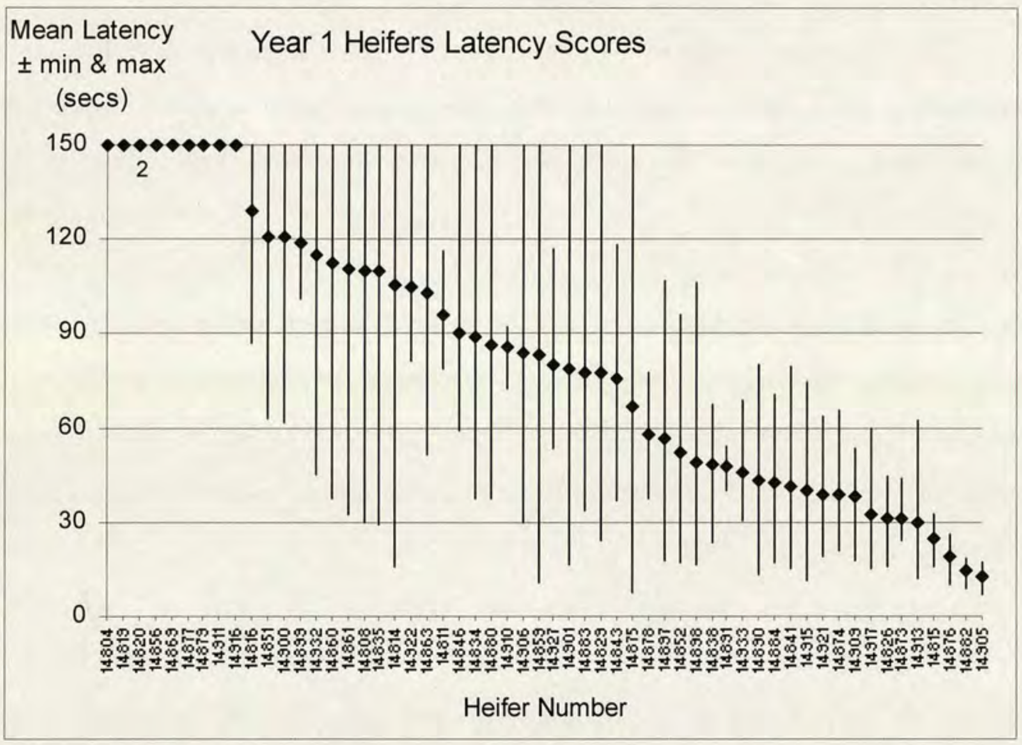


Fig. 3.12: Latency scores from the Handling Tests carried out on 56 Year 1 heifers. Latency = Time taken for the animal to move into the test corner for 30 seconds. The means are from three tests unless indicated by '2'.

3.3.4.4. HA Test Results Summary

In summary, the HA test was carried out three times on a large number of bull and heifer calves. After preliminary comparisons between two methods of analysis, examining Latencies and calculating a Docility Score, the simpler method of looking at latencies and durations of individual behaviours was chosen for the full analysis. This decision was taken as the Latency measure showed a wider range than the Docility Score, and was a more repeatable measure, in the heifers at least. Furthermore, Latency was very highly correlated with the Docility Score in both the bulls and the heifers. This implied that it was the main variable (or was highly correlated to other variables) that determined the Docility Score. It appears that using the Latency measure is a simple alternative. Further, the Docility Score combines a range of variables that may not all be reflective of the same underlying motivations, and assigns a somewhat arbitrary value to each, seemingly without prior investigation of the relationships between the different variables.

The median Latency across all the bulls did not change between the test repeats, but decreased in the repeat tests carried out on the heifers. There was no relationship seen in latency measures between Test 1 and successive tests in the bulls, but there was a small relationship between Tests 2 and 3. Repeatability of Latency was much higher in the heifers, the highest being a marked relationship between tests 2 and 3.

'Percentage of time standing in corner' was highly correlated to the Latency measure. This measure was not repeatable between bull tests, but showed a moderate to high relationship between heifer tests. 'Percentage of time running or walking' was not repeatable between bull tests, but showed substantial relationships between heifer tests.

3.4. Discussion

3.4.1. Flight-from-Feeder Test

The FF scores obtained in the tests showed both a wide variability between the animals, and a high level of repeatability (≥ 0.5) when the test was repeated on the same animals, for both the bulls and the heifers. The repeatability coefficients over the three tests of 0.52 ± 0.09 for the bulls and 0.50 ± 0.08 for the heifers mean that at least 50 % of the variation seen between test repeats were due to consistent individual responses. This test is therefore a useful measurement of the behaviour of both the bull and heifer calves.

Some differences in responses were seen between the bulls and heifers. The heifers were generally tamer than the bull calves, 32 % allowing themselves to be repeatedly approached and touched, gaining a maximum score of 6. This is probably due to the close human contact they received during artificial rearing, whereas the bull calves were left with their dams in the field for six months before weaning and therefore had less contact with humans. It is also possible that the heifers appeared less fearful as they were likely to have been experiencing higher motivation than the bulls to stay at the feeders, due to their restricted access to food.

Although the heifers were tamer than the bull calves, they showed a decrease in score (i.e. became more fearful) between Tests 1 and 2, that did not occur with the bull calves. The highest repeatability coefficient obtained by the heifers was also between Tests 1 and 2, suggesting that the decrease in response was consistent across the whole group. The novelty of the test situation may have contributed a large part to the response seen when the test was carried out for the first time, and possibly the motivation of the heifers to investigate the human observer was stronger than during the subsequent tests. Perhaps the most likely explanation is that the heifers were very accustomed to farm staff walking close to the feeders, and walking past without approaching or touching them. It is likely that they allowed the observer to approach initially because they did not expect her to come so close or to attempt to touch them.

Becoming more wary of the observer in the second test would lead to the decrease in scores seen. It is also likely that the bulls were more wary of the observer prior to the first test, as they had not entered the pens in the shed until the age of six months, after being at pasture with their dams, and so may not have habituated to people walking past them to the same extent as the heifers.

The highest repeatability coefficient obtained by the bulls was between Tests 2 and 3 in the bulls, also implying that there were factors affecting scores in Test 1 that did not affect the second and third tests. The results from Tests 2 and 3 may therefore give the most accurate measure of fearfulness of human approach, with the novelty of the test situation apparently playing a much smaller part in the responses to the repeat tests, in both bulls and heifers.

This test was designed so that each animal did not need to be separated from its penmates in order to be tested. On the other hand, the animal was visually restricted while it was standing in a feeder, and so its response was unlikely to have been influenced by the behaviour of other animals in the pen. This test situation provided a good compromise between approaching each animal in the pen, where the proximity of herdmates may have affected their responses, and taking the animals out of the pen one at a time to test them under more standard social conditions, which would have added an additional factor of social separation. The results found in a previous study indicate that separation from conspecifics can influence an individual's responses to flight tests. Fisher and his colleagues (2000) carried out a yard flight test on Limousine x Jersey cross animals. Each animal was placed next to a pen of six or more of its penmates, and approached by an observer from a distance. The test was repeated three times at monthly intervals, and showed a similar level of repeatability as this present study, of 0.51 ± 0.03. However, the test scores correlated more strongly with a sociality test carried out at the same time, than they did with another flight test which was carried out in a paddock. The sociality test measured the time taken for each animal to move down a yard to return to its conspecifics when separated. One explanation of the stronger correlation between the yard flight test and the sociality test may be that the social environment of the animals during the flight test played a larger part in the responses of the animals than the effect of human approach.

In conclusion, the FF test produces results of high variability between calves, and high repeatability, and therefore is useful as a measure of fearfulness of human approach in cattle.

3.4.2. Social Separation Test

The SS Test showed a wide range of response between individual bulls and heifers. The combined duration of Walk, Escape and Run (WER), thought to be a measure of sociality in this test situation, showed high intra-animal repeatability in both sexes. The repeatability of WER exceeded 0.5 across all three tests and across both pairs of tests examined in the heifers, and across all three tests and across Tests 2 and 3 in the bulls. This test therefore provides a useful temperament measure in both the bulls and the heifers.

The highest repeatability was seen between Tests 2 and 3 in both the bulls and the heifers. No overall change in test score was seen between repeated tests on the bulls, but an increase in the scores was seen in the third test repeat on the heifers, implying that their levels of sociality motivation were higher during the final test. The high repeatability seen between the second and third tests implies that this increase was consistent across the heifers. The increase in score implies the heifers did not habituate to the test procedure over the three days, but instead became more fearful as they learnt the procedure. This, along with the extreme responses seen by a few animals that tried to jump over the gates, demonstrates that this test caused high levels of fearfulness in some of the animals.

The heifers also showed generally higher WER scores than the bulls. They may have experienced higher motivation to rejoin their penmates as they would have become much more accustomed to being in close contact with each other than the bulls. The heifers were housed in groups from the age of two weeks. In contrast, the bulls were reared at pasture with their dams, and not grouped together in pens until they were six months of age. This difference in rearing systems may explain why the heifers appeared to be more closely bonded to their penmates, and experienced higher levels of sociality than the bulls.

The higher repeatability seen between Tests 2 and 3, compared to between Tests 1 and 2, in both sexes implies that the novelty of the situation may have played a role in the animals' responses to the first test that it did not play in responses to the later tests. Therefore the best measures of the single trait of sociality may be obtained by using the scores from the latter tests.

The repeatability estimates obtained for this test were similar to repeatability values reported by Hopster and Blokhuis (1993), when they isolated Holstein cows by keeping them inside a cubicle house after milking while the rest of the herd went out to pasture. They reported estimates of 0.93, 0.69 and 0.69 for numbers of vocalisations, number of visits to an exit door and time spent pacing, respectively. The estimates obtained here were higher than those obtained by the Jersey x Limousin cattle in the sociality test described in Section 3.4.1 (Fisher *et al.*, 2000). These animals were separated from their herdmates and the time taken for them to move down a yard to return to them was measured. The test was conducted three times at monthly intervals. A moderate repeatability of 0.34 ± 0.04 was obtained. A similar test was carried out in this study, in an experiment described in Chapter Four, in order to validate the SS Test. The results obtained by Fisher are discussed further there.

To conclude, the SS Test produced results that showed high variation between animals, and high repeatability within animals. It is therefore a useful test of a temperament trait, probably sociality, in cattle. The results of the test are examined further in Chapter Four, where an experiment carried out to validate whether the SS test measures sociality is described.

3.4.3. Novel Object Test

Very little variation between individual responses to the traffic cone was seen between the Year 1 bulls and the Year 1 heifers. The cone did not induce the expected levels of fearfulness in the animals, and the majority approached it with little or no hesitation. When the test was carried out twice on the Year 2 heifers, a plastic barrel was chosen for the second test, in the hope that the larger object might cause fear in some of the animals. However, the latencies to approach the barrel were just as short as those for the cone.

Similar objects have been used successfully in several studies of neophobia in cattle. Novel objects used have included colourful striped inflatable balls (Murphey *et al.*, 1981; Hemsworth *et al.*, 1996), an iron truncated pyramid with white and green stripes (Boissy & Bouissou, 1995) and a white plastic bucket and a white chair (Hemsworth *et al.*, 1996). Boissy *et al.* (1998) reported large individual variation in the latency of Aubrac heifers to approach a traffic cone (although in their study the cone was sprayed with urine, which may have contributed to the longer latencies!). The animals in the present study were intensively-housed at the time of testing, and were likely to have seen new objects regularly, when being fed, or when being moved to different areas of the shed for weighing, for example.

Variation was seen in the amount of time the animals spent exploring the novel object. Some of the animals spent a large percentage of the test time licking, pushing or rubbing the cone, while others touched and licked it briefly and then moved away. Hence it was believed that the test may have been of use in measuring investigatory motivation. However, the repeatability of this measure in the Year 2 heifers was found to be moderate at 0.26. Although this signified a substantial relationship, it is not a strong enough relationship to be confident that this is a good trait measure.

It was therefore concluded that the NO Test was of little use in investigating fearfulness or investigatory traits in these animals.

3.4.4. Handling Test

The main measure used in the HA Test, Latency, showed considerable inter-animal variation for both the bulls and the heifers. The measure showed poor intra-animal repeatability in the bulls, but moderate to high repeatability in the heifers. Repeatability of Latency between Tests 2 and 3 in the heifers (0.53 ± 0.97) exceeded the minimum cut-off value of 0.5, and repeatability between Tests 1 and 2 almost reached this value (0.47 ± 0.11). This study therefore confirmed that the Handling (or Docility) Test can be a useful method of assessing a trait in artificially-reared heifers that relates to the ease of handling of the animal. This trait is most likely fearfulness of humans. This is also the first time the test has been used with bull calves.

Why differences in response were seen between the bulls and the heifers is not clear. All the reported studies using the test have used young heifers (Boivin *et al.*, 1992a; Le Neindre *et al.*, 1995; Grignard *et al.*, 2001) except possibly one in which the gender of the calves is not reported (Grignard *et al.*, 2000). Between-sex comparisons of behavioural results on these animals will always be complicated by the differences in the rearing systems used for the bulls and the heifers, and there are two additional factors that may have contributed to the difference in repeatability of the test results. Firstly, the HA Test was carried out before the other behaviour tests with the bulls, but after the other tests in the heifers. Secondly, a gap of two days was left in between each test for the bulls, but the heifer tests were carried out on successive days. Differences in exposure to humans during rearing, and differing experience of other test procedures prior to testing may be expected to lead to differences in latencies obtained, but why these factors should contribute to a difference in repeatability is not clear. A gap of two days between tests would not be expected to have a large effect, especially since a moderate correlation was found between tests carried out seven months apart on a group of Limousin heifers (Grignard *et al.*, 2001). As this was the first time the handler had carried out the test, her behaviour towards the bulls during the test may not have been as consistent as it was for the heifers later on, when the tests had been well practiced. Additionally, the bulls were fairly well grown, and one at least started displaying aggressive behaviour.

Another interesting finding was the higher repeatability seen between Tests 2 and 3 than between Tests 1 and 2, in both the bulls and heifers. In the heifers a decrease in median Latency was seen in Tests 2 and 3 from Test 1. It appears that the novelty of this situation played a significant part of the animals' responses during the first test, and levels of fear decreased in subsequent tests as they habituated to the test procedure, allowing the handler to get the animal into the corner in a shorter time. The measure of Latency in this test was assumed to represent a single motivatory state of fear of humans therefore, it may be more useful to take the measurement from Tests 2 and 3 and not include the measurement from Test 1.

Along with the novelty of the situation, social context may play a part in the animals' responses to handling (Grignard *et al.*, 2001). Grignard and colleagues compared cattle reactions to the test with or without visual contact with familiar peers. During handling, the human required more time to restrain the calf in the corner when the conspecifics were present. The effect of novel surroundings is also a potent condition that causes fear. In order to minimise these factors in this study, the tests were carried out in the animals' home pens with their herdmates in one neighbouring pen and the usual neighbouring group in the other. Therefore the Latencies seen in this study may represent more accurate measurements of fearfulness of humans than previous tests that were carried out in pens unfamiliar to the animals (Boivin *et al*., 1992a; Le Neindre *et al*., 1995; Grignard *et al*., 2000; Grignard *et al*., 2001). The handler has also been found to have a significant effect on the test scores (Le Neindre *et al*., 1995). A single handler was used for all the testing in this study.

In conclusion, the HA Test was found to give useful measures of the docility of the cattle. The Latencies from the heifers and the bulls showed high inter-animal variation. The heifers' scores also showed high intra-animal repeatability. The test scores showed poor repeatability in the bulls, probably due to methodological differences.

3.5. Conclusions

In this study four behavioural tests were assessed according to two criteria, inter-individual variation, and intra-individual repeatability. Data from two of the tests, the FF and SS Tests, showed wide variation and high repeatability in both sexes. The NO Test showed low variation between animals of both sexes in the main response measure. A possible alternative measure was investigated that showed high variation in response, but only moderate repeatability. The HA Test showed wide variation between animals of both sexes, with high repeatability in the heifers but not in the bulls. Possible methodological reasons for this have been discussed. The FF, SS and HA were therefore judged to be potentially useful measures of fearfulness traits, and were used in further experiments to investigate fearfulness in cattle.

As the repeatability of responses to the FF and SS Tests (and the HA in the heifers) was shown to be high, the number of repeats of the tests carried out on the animals born in Year 2 and 3 was decreased to two. Repeatability estimates between test repeats were found to be the highest between the second and third test repeats in the FF Test (both sexes), the SS Test (both sexes) and the HA Test (heifers only). This is thought to be due to the novelty of the test situations having a large effect on the responses to the first test. Therefore it was considered appropriate to use each animal's response to the *second* repeat of each test as the trait measure to be used in further comparisons or analysis. Each test was designed to measure behaviour resulting from one main motivation. It appeared the first repeat of each test also incorporates novelty as an additional, unwanted, component.

Test measures can be validated by demonstrating that the responses seen relate to behaviour seen in another similar situation (see Chapter 1.5.1). The measures from the FF, SS and HA Tests were examined further, to see if they could be validated as fearfulness trait measures, and these experiments are described in Chapter Four. The present study examined the repeatability of the test results over just a few days. It remains to be demonstrated that these tests measure aspects of temperament that are consistent over some time. Hence, Chapter Five looks at the consistency of the test measures taken over longer periods of time, and whether they are useful to predict future behaviour in young animals. Chapter Six investigates the genetic basis of two of the traits.

Chapter Four: Validation and Interpretation of Behavioural Test Results

4.1. Introduction

This chapter is concerned with the issue of validation of behavioural tests. As described in Chapter 1.5.1, validity concerns the extent to which behavioural measurements actually measure the underlying temperament traits the investigator wishes to measure (Martin & Bateson, 1993). Even if the responses of individual animals in a test situation are shown to be repeatable, and hence seem to be reflective of a trait, the responses may be very specific to aspects of that particular test situation (Mendl & Harcourt, 1988). The test responses may not give useful information about how an animal will respond to similar stimuli in different situations. Therefore, interpretation of responses from behavioural tests is greatly enhanced if they can be shown to correlate with behaviours seen in real-life situations (see Section 1.5.1.2). Looking for relationships between different test situations provides a step towards validating the measures taken from each. By necessity, the interpretation and validation of responses elicited by a test are often carried out retrospectively.

This chapter examines whether the scores measured in the tests carried out on the Year 1 bulls and heifers, that were described in Chapter Three, can be validated by comparing results from different tests that use similar fear-inducing stimuli. The chapter is divided into two parts. Firstly, the results from three of the tests were examined further to investigate how measures taken in one test situation related to measures taken in the other tests, in order to validate the Flight-from-Feeder (FF) Test and the Handling (HA) Test (Section 4.2). Secondly, an additional experiment was carried out in order to validate the Social Separation (SS) Test (Section 4.3).

The FF and HA Tests were both thought to measure fearfulness of humans. Previous attempts have been made to relate the behaviours seen in these types of tests to behaviours seen in other situations. The relationship between flight distance and approach distance recorded from approach tests has been examined. Approach tests measure the behaviour of an animal when put in close proximity to a passive, stationary observer for a set amount of time. A high correlation (0.71) was found between flight distance and 'speed of movement' in an approach test, in a study looking at various temperament test measures in six different breed groups (Fordyce *et al.*, 1982). Murphey *et al.* (1981) studied the relationship between flight distance and approach distance in herds representing several *Bos taurus* and *Bos indicus* breeds. The 21 herds in the study were ranked according to flight distance and approach distance, and these two variables were moderately positively correlated ($r = 0.49$). These two studies examine differences between groups rather than differences between individuals. Purcell *et al.* (1988) found correlations between flight distance and approach distance in Holstein dairy cows, both between *and within* herds, but these values were low ($r = 0.18$ within herd). Considered together, these studies show there is some, but not strong, evidence that flight test responses are related to approach distances, but the relationships seen are not strong enough to provide validation.

Further evidence that flight distance relates to fearfulness of humans is provided by the finding that the measure varied between groups of Simmental x Angus steers that had been given different prior levels of handling (Matthews *et al.*, 1997). A similar study was carried out in Friesian calves (Krohn *et al.*, 2001). Three groups of calves handled at different times after birth all showed a shorter flight distance compared to the control group that received no handling. The authors concluded that the flight distance reflects an animal's fear of humans, as it discriminates handled from non-handled calves. Again, these experiments looked at group differences, rather than differences between individuals. So although these experiments showed that flight distance is a meaningful measure that is useful for discriminating between different treatment groups, it has not been validated by correlation of variables measured from individual animals.

Evidence that handling (or Docility) tests measure fearfulness of humans also comes from the differences in test scores seen between groups of heifers reared in management systems which incorporated different levels of human contact (Boivin *et al.*, 1992a). Salers calves from a ranging system, which had little contact with handlers, took longer to be brought into the corner and to remain there for 30 seconds than those reared in an indoor system which allowed them to suckle from their mothers twice a day under human control. Le Neindre and his colleagues (1995) tested hundreds of Limousin heifers that were sired by 34 bulls but reared on many different farms. The husbandry system again had an effect on Docility Score, with heifers managed indoors being more docile than those reared out-of-doors. Although the test has been shown to measure fearfulness of humans in this way, it again has not been validated as a test suitable for distinguishing differences between individuals within a herd.

In this study, relationships between measures taken from the FF, SS and HA Tests carried out on the bulls and heifers from Year 1, and described in Chapter Three, were examined. The main aim was to determine whether a relationship was seen between the FF and HA Test measures that would validate both these test procedures. No previous studies have examined relationships between results from these two tests. Further aims were to examine relationships between other test variables to aid the interpretation of the behavioural traits. The objectives of the analysis are summarised in the following five hypotheses. The first three of the hypotheses are concerned with the main measures from each test and how they relate to one another. The latter two hypotheses use other behavioural measures from the tests to aid interpretation of the test results.

1. The Flight Scores from the FF Test and 'Latency to stand in the test corner for 30 seconds' (Latency) from the HA Test were both thought to be measures of fearfulness of humans. Therefore a relationship was expected between these measures. If a strong relationship was seen, this would validate both variables as measures of fearfulness of humans.
2. The Flight Scores and the duration of time spent Walking, showing Escape behaviour or Running (WER) in the SS Test were thought to measure different traits. The Flight Scores were thought to measure fearfulness of humans and the WER measure was thought to reflect fearfulness when socially isolated. These are thought to be different traits, and no relationship between the two variables was expected.
3. Again, Latency from the HA Test and WER from the SS Test were thought to measure different traits, and no relationship between these two variables was expected.

4. The level of separation from conspecifics was designed to be much greater in the SS Test than in the HA Test. However, separation from penmates was experienced by the cattle in the HA Test situation as well, but it was hoped that this was minimal and that it would not impact upon the response to the test. To test this, levels of WER seen in the first 30-second 'before handling' period of the HA Test (for details of the test procedure, see Chapter 2.3.5) were compared with levels of WER seen in the first 30 seconds of the SS Test. Much higher levels of WER were expected in the SS Test.
5. Flight Tests are widely believed to measure fearfulness of humans (e.g. Matthews *et al.*, 1997). However, different levels of investigatory motivation may also play a part in the animals' responses. The FF Scores were therefore compared to another potential measure of investigatory motivation from the SS Test - 'time spent in contact with the wooden boards'. It was believed that fearfulness of humans was the dominating motivation in the FF Test and a relationship between these two variables was not expected.

The second part of this chapter describes an experiment that aimed to validate the SS Test using another test that was also thought to measure sociality. Tests where social separation is imposed while the animal is held in a familiar environment have seldom been used in cattle (Hopster & Blokhuis, 1993; Boissy & Le Neindre 1997), and no attempt to validate the interpretation of behavioural responses displayed in such a situation has been made. Tests of sociality are commonly used in poultry (for example, Mills *et al.*, 1993; 1995; Hocking *et al.*, 2001; Marin *et al.*, 2001; Jones *et al.*, 2002). One method is a runway test, in which each bird in turn is placed in a runway with a goal box containing conspecifics at the opposite end. Measures such as the latency of the bird to enter the zone near the conspecifics are recorded. Such a test was described by Jones *et al.* (2002), and the Sociality (SO) Test carried out on the cattle for the experiment in Section 4.3 was based on this test. It was designed to measure the animals' responses to another situation involving separation from the herd. This allowed the testing of a sixth hypothesis:

6. The SS Test and SO Test are both thought to measure sociality. Therefore a relationship between scores from the two tests was expected. A strong relationship would validate measures taken in both tests.

4.2. Relationships between Test Measures

4.2.1. Methods

The analysis in this chapter was carried out using results from the FF, HA and SS Tests carried out on the 49 bulls and 54 heifers from Year 1, which were described in Chapter Three. The test procedures are described in Chapter 2.3, and the measures taken discussed in Chapter 3.3. Briefly, the FF Test was thought to measure fearfulness of humans, and the measure taken was a Flight Score ranging from 1 - 6. The SS Test was thought to measure sociality, and the main measure taken was the total duration of the test time spent Walking, showing Escape behaviour and Running (WER; see Chapter 3.3.2.1). The HA Test was thought to measure fearfulness of humans, and the main measure taken in this test was 'Latency to stand in the test corner for 30 seconds'. The results from each of these tests were found to be repeatable and show variation between the animals (Chapter 3.3.3).

In Chapter 3.5, it was concluded that measures taken in the second repeat of each test were the most useful for reflecting the temperament traits of interest. The novelty of each test situation appeared to have a substantial effect on the animals' responses to the first run of each test. Measures from the second repeat of each test were therefore used for the analysis in this section. The measures used and their medians and ranges are summarised in Table 4.1. Pairs of scores from the tests were plotted and correlations calculated, to look for evidence of relationships between individual's measures from different tests.

As most of the data were not normally distributed, and could not be transformed, Spearman Rank correlations were used throughout. Differences in durations of behaviours between different tests were examined using Paired t-tests.

Table 4.1: Summary statistics of the behaviour measures that are analysed in Chapter 4.2. The data from the *second* repeat of each test are used.

Test	Measure	Heifers				Bulls			
		n	min	max	median	n	min	max	median
FF2	FF Score	53	1	6	5	48	1	6	3
HA2	Latency	55	7	150	58	48	11	150	78
	WER1st30	55	0	10	2	48	0	12	0.5
SS2	WER	54	0	250	39	43	0	217	31
	WER1st30	54	0	24	5	43	0	23	3
	B	54	0	216	82	43	0	280	111

The three tests and six measures are:
FF = Flight-from-Feeder Test, FF Score = Flight-from-Feeder Scores;
HA = Handling Test, Latency = Latency to stand in the test corner for 30 seconds, WER1st30 = duration of Walking, Escaping and Running seen in first 30 seconds;
SS = Social Separation Test, WER = duration of test time spent Walking, Escaping and Running, WER1st30 = duration of Walking, Escaping and Running seen in first 30 seconds, B = duration of test time spent sniffing, licking or rubbing the wooden boards.

4.2.2. Results and Discussion

The results are described and discussed according to the five hypotheses.

1. Is there a relationship between the Flight Scores from the FF Test, and Latency from the HA Test?

Spearman Rank correlation coefficients were calculated between Flight Scores and Latency. A low FF Score was thought to indicate high fearfulness of humans in the FF Test, and a high Latency was thought to indicate high fearfulness of humans on the HA Test. No relationship was seen between the two measures in the bulls ($n = 48$, $r_s = -0.09$, $p = 0.557$; see Fig. 4.1). A small positive correlation was found between the two measures in the heifers ($n = 50$, $r_s = 0.28$, $p = 0.045$; see Fig. 4.2), indicating that heifers that were the most approachable in the FF Test showed longer latencies to be held in the corner in the HA Test. The lack of a relationship between the two measures in the bulls implies that the two tests do not measure the same trait.

The correlation between the two test measures in the heifers implies that there is a common element to both tests. Both tests are thought to measure fearfulness of humans. However, the correlation is positive. A *long* latency in the HA Test (that was assumed to signify *high* fearfulness) is associated with a *high* FF Score, which is assumed to indicate *low* fearfulness. Although the correlation between the two measures is low, this disparity between the results from the tests implies that one (or both) cannot be a good measure of fearfulness of humans, as carried out here. The most likely explanation may be that additional factors other than fearfulness of humans, such as aggressiveness, affect responses to the HA Test. Aggressiveness towards humans may play a larger part of the response to the HA Test than the FF Test because the contact between the animals and the human handler is closer. In the FF Test, the behavioural responses available to the animal are limited - to stay still while the human approaches or to move away. Because the animal and the observer are physically separated by the feeder (see Fig. 2.5) there is not much opportunity to exhibit aggressive behaviour. The HA Test situation, in contrast, gives fearful animals the opportunity to react in different ways. Many fearful animals had a flight zone of a few metres and ran around the pen as soon as approached, and could not be held in the corner. Some fearful animals were able to be held in the corner, and stood with their heads down and their bodies tense. The animals that did not appear fearful also differed in their responses. Some animals that could be easily approached and touched walked into the corner easily and appeared relaxed and turned to sniff or lick the handler. Others that were easily approached refused to move, or moved slowly around the pen and would push past the handler to avoid being held in the corner.

Additionally, it is possible that responses in the FF Test were affected by the level of food motivation of the heifers due to their restricted feeding regime, which is unlikely to have played a part in responses to the HA Test. It is also possible that separation from penmates for the HA Test, although to a much lesser degree than occurred in the SS Test, contributed to the responsiveness seen in the HA Test. This is investigated in Hypothesis 4. Hence this test may not be the best method with which to measure the single trait of fearfulness towards humans.

It is surprising nonetheless that such a lack of correlation is seen between the two tests. Both flight tests and handling tests have been shown to be moderately correlated with crush tests in two separate studies (Andrade *et al.*, 2001; Grignard *et al.*, 2001). During crush tests the cattle were restrained in a squeeze chute and their behaviour recorded or assessed according to a five-point system. Andrade *et al.* (2001) carried out the flight test two months after a temperament assessment was made of 18 Brahman cows in a crush, and found a moderate correlation of 0.6 between the measures. Grignard *et al.* (2001) compared the responses of Limousin heifers tested in a handling test at nine months of age and in a crush test at 12 months. The animals' behaviour in the crush was measured twice, once with and once without the presence of a human. A small relationship (Pearson correlation, $r = 0.29$) was seen between the handling and crush test when no human was present, and a moderate relationship ($r = 0.37$) was seen when a motionless man stood beside the crush. The size of the correlations reported in these two studies are higher than the correlations found in this study between the FF and HA Tests. Such strong relationships were surprising as it may be expected that extra factors which were not present in the flight or handling test situation, such as physical restraint, might have affected the responses of the animals to the crush test.

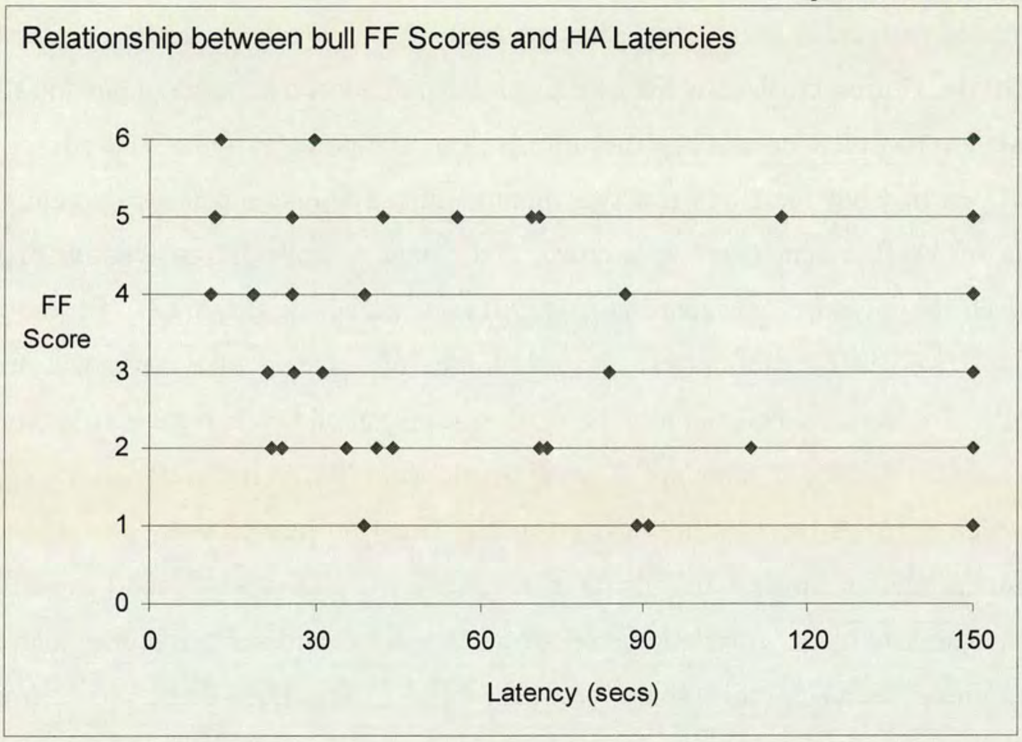


Figure 4.1: Relationship between Flight Scores from the second Flight-from-Feeder Test and 'Latency to stand in the test corner for 30 seconds' from the second Handling Test carried out on 48 Year 1 bulls.

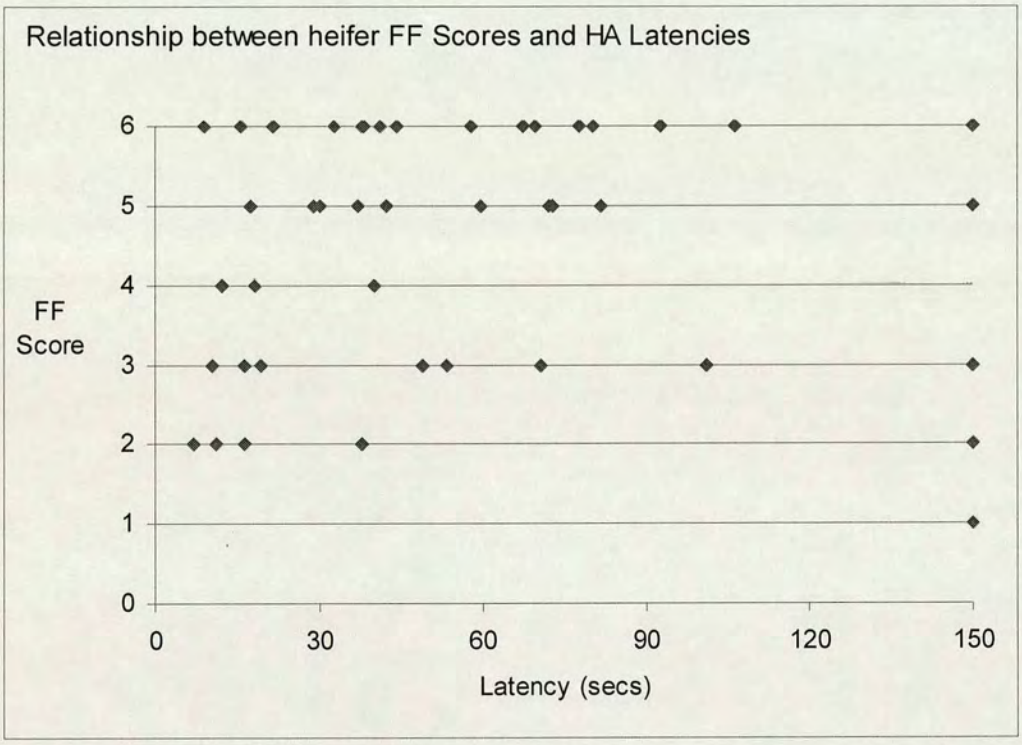


Figure 4.2: Relationship between Flight Scores from the second Flight-from-Feeder Test and 'Latency to stand in the test corner for 30 seconds' from the second Handling Test carried out on 50 Year 1 heifers.

2. Is there a relationship between the Flight Scores (thought to measure fearfulness of humans) and the WER Scores from the SS Test (thought to measure fearfulness due to social isolation)?

A tendency towards a small negative relationship between these scores was seen in the bulls (Spearman Rank correlation; n = 43, r_s = -0.28, p = 0.066; Fig. 4.3), indicating that tamer bulls in the FF Test tend towards less fearful behaviour in the SS Test. No relationship between the scores was seen in the heifers (n = 51, r_s = -0.16, p = 0.251; Fig. 4.4). The result in the heifers supports the hypothesis that the two tests are measuring different traits. There is no evidence that the factors that are common to both the test procedures, such as the presence of humans, are confounding the results from the tests. The tendency towards a relationship in the bulls suggests that the handling by the humans to separate the animal from the group and put it into the test pen may impact on the responses to the SS Test. Handling may have affected the bulls to a greater degree than the heifers, because they were generally less tame due to their different rearing experiences. However the effect is so small as to be negligible (Fig. 4.2), and the two measures can be considered independent measures of two different traits in both sexes.

Two previous studies have reported a relationship between scores from flight tests and tests of social motivation (Matthews *et al.*, 1997; Fisher *et al.*, 2000). In both, it seems likely that the relationships seen were due to the common factor of human presence that had an impact on both tests. Matthews *et al.* (1997) suggested a link between flight distance and a 'sociability' measure. They found that intensive handling of Simmental x Angus steers during rearing reduced both flight distance, and the strength of attraction of an isolated animal to its familiar herdmates. It is possible that the role that humans played in the set-up of the sociability test influenced and confounded the results of the test. As this reference is a conference abstract, few details are available.

Fisher *et al.* (2000) found a relationship between scores from a flight test and a sociability test in Jersey x Limousin cross heifers and steers. The sociability test was designed to measure the motivation of an animal to be with its herdmates if separated. A group of six or more cattle from the test group were held in a pen at the bottom of a handling yard. The remaining animals were released one at a time at the top of the yard 30 m away, and time taken for each animal to move down the yard to join it's conspecifics was recorded. The flight distance was recorded in the same yard, approximately 30 seconds after the completion of the sociability test, when an animal was placed at the bottom of the yard close to the pen containing its herdmates. A person walked slowly towards the stationary animal. A substantial relationship was seen between the scores from the two tests (correlation = -0.44). The two tests were carried out in such quick succession that the relationship seen between them is likely to be due to a common factor that affected both the tests. This may either be fearfulness of humans or separation from herdmates.

It seems likely, from the results of the present study and the two other studies described here, that fearfulness of humans and sociality can be treated and measured as two separate traits. Any small associations found between tests can be explained by the influence of additional factors acting on both the test situations.

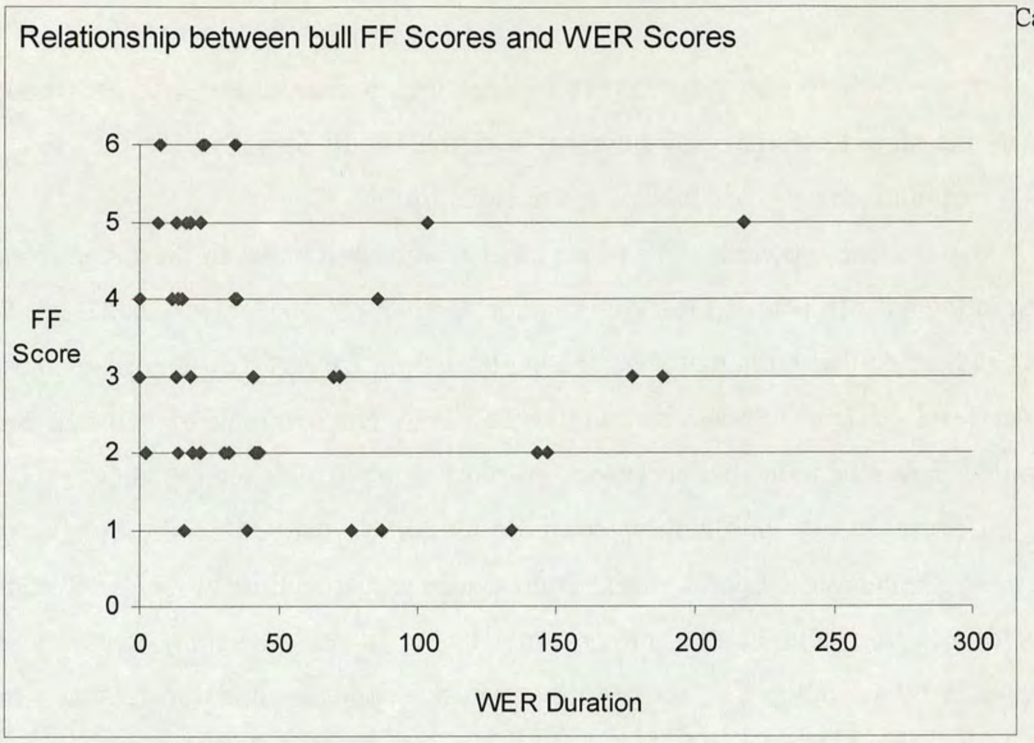


Figure 4.3: Relationship between Flight Scores from the second Flight-from-Feeder Test and durations of WER (Walk, Escape and Run) from the second Separation Test carried out on 43 Year 1 bulls.

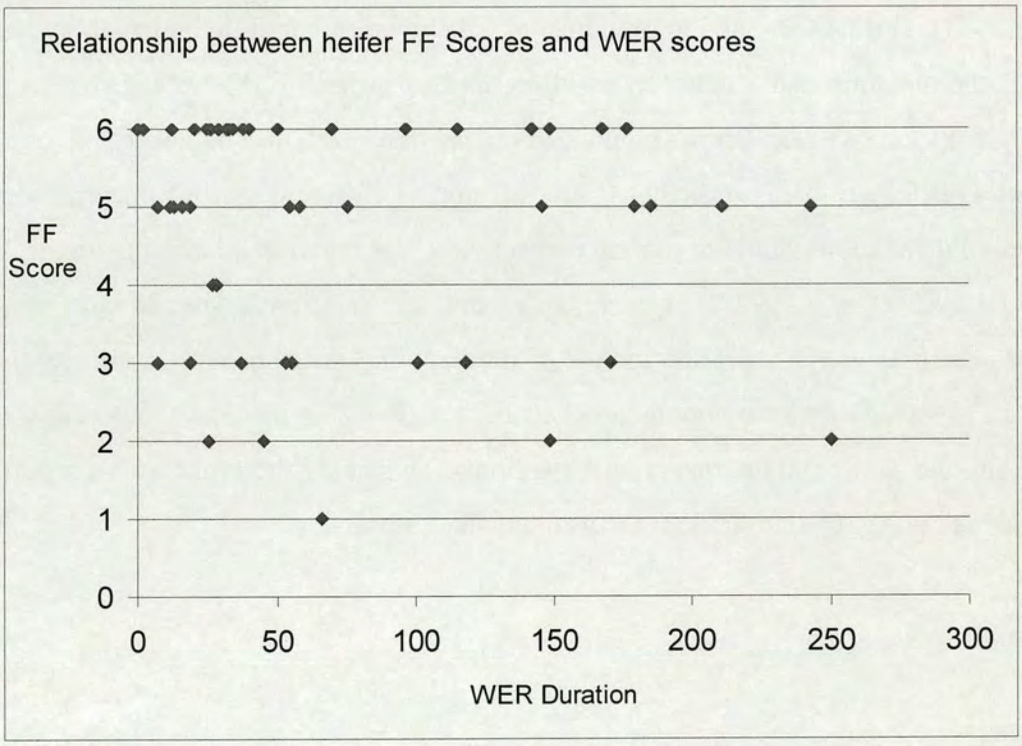


Figure 4.4: Relationship between Flight Scores from the second Flight-from-Feeder Test and durations of WER (Walk, Escape and Run) from the second Separation Test carried out on 51 Year 1 heifers.

3. Is there a relationship between the Latency measures from the HA Test (thought to measure fearfulness of humans) and the WER Scores from the SS Test (thought to measure fearfulness due to isolation)?

A tendency towards a small negative relationship between these variables is seen in the bulls (Spearman Rank correlation; $n = 43$, $r_s = -0.298$, $p = 0.053$; see Fig. 4.5), indicating that bulls that show a long latency in the HA Test tended to show a lower level of fearful behaviour in the SS Test. No relationship between these variables was seen in the heifers ($n = 53$, $r_s = 0.133$, $p = 0.343$; see Fig. 4.6).

These results are similar to those described above for the relationship between FF and WER Scores. There is no evidence that additional factors confound the testing procedure in the heifers, but slight evidence they play a part in the responses of the bulls. The relationship is in the opposite direction to that which would be expected if fearfulness of humans were affecting the SS Test measures. This presumes that a long latency in the HA Test signifies high fearfulness, and as discussed in point 1, it does not seem that this is the case. As can be seen from Fig. 4.3, any relationship is so small as to be negligible.

The closest comparison previously made between handling and sociality tests were comparisons made between handling and open-field (OF) tests. OF tests are thought to impose fear from several stimuli, as described in Chapter 1.3.4, one of which is isolation from herdmates. Calves from two different rearing systems which imposed different amounts of human contact were compared in a handling and an OF test (Boivin *et al.*, 1992a). A difference between traditionally-reared and range-reared animals was seen in the results of the handling test, but not the results from the OF test. The authors concluded that the handling test and the OF test did not measure the same characteristics of the animals. This study therefore adds support to the idea that the HA and SS Tests measure different traits.

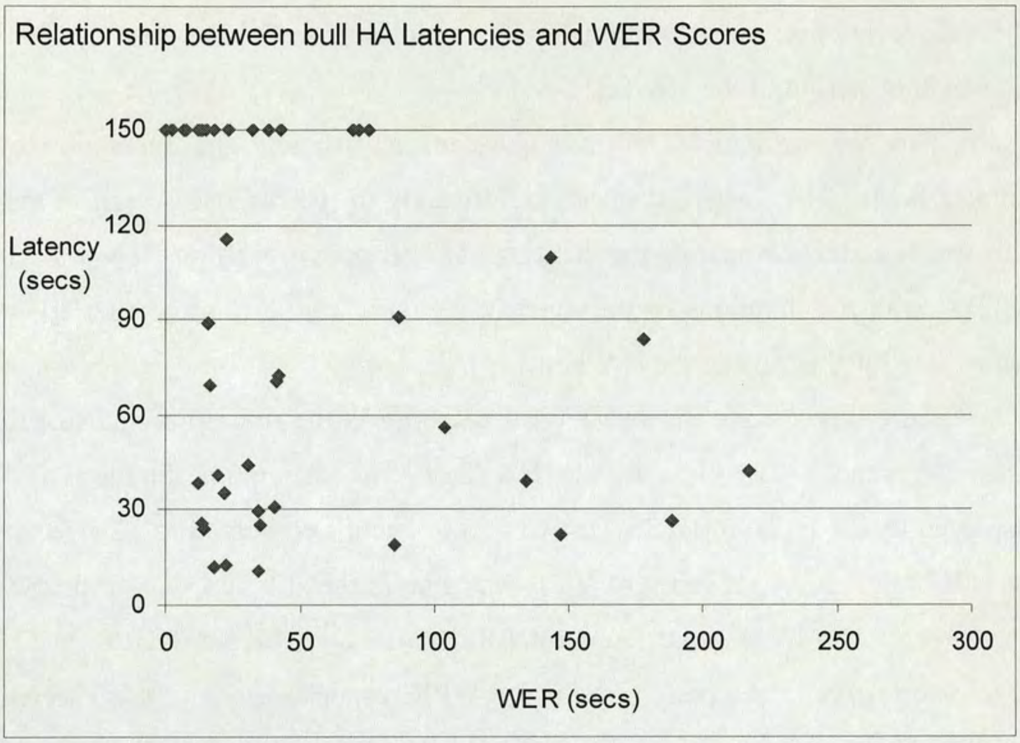


Figure 4.5: Relationship between 'Latency to stand in test corner' from the second Handling Test and durations of WER (Walk, Escape and Run) from the second Social Separation Test carried out on 43 Year 1 bulls.

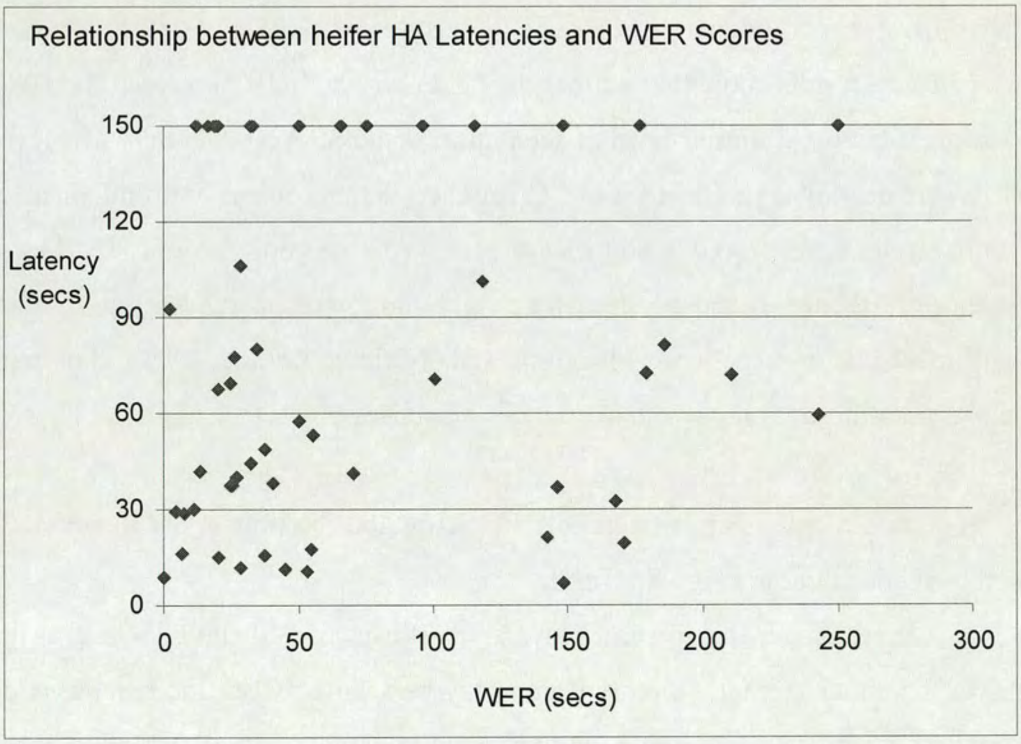


Figure 4.6: Relationship between 'Latency to stand in test corner' from the second Handling Test and durations of WER (Walk, Escape and Run) from the second Social Separation Test carried out on 53 Year 1 heifers.

4. Is the level of response to separation greater during the SS Test than the first 'pre-handling' period of the HA Test?

This was examined to investigate further whether the separation of the animals for the HA Test contributed significantly to the responses seen in the test. This was tested by comparing the duration of WER seen in the first 30 seconds of the SS Test with the durations of the same behaviours seen during the first 30-second 'before handling' period of the HA Test.

Thirty-eight bulls showed a WER response in the first 30 seconds of the SS Test, compared to 35 bulls in the HA Test. The maximum duration of WER displayed by the bulls in the SS Test was 23.0 seconds compared to 12.1 seconds in the HA Test. A lower duration of WER was seen in the HA Test than in the SS Test (Paired t-test; $n = 45$, mean difference = 1.8 seconds, $t = 2.16$, $p = 0.036$).

Forty-five of the heifers showed a WER response in the first 30 seconds of the SS Test, compared to 31 heifers in the HA Test. The maximum duration of WER displayed by the heifers in the SS Test was 24 seconds compared to 9.9 seconds in the HA Test. A lower duration of WER was seen in the HA Test than in the SS Test (Paired t-test; $n = 55$, mean difference = 3.1 seconds, $t = 3.95$, $p < 0.001$).

These results show that neither the heifers or the bulls perceived the HA Test situation as having a similar level of separation as the SS Test, and only low levels of WER were displayed in the HA Test. This indicates that separation from conspecifics is unlikely to have played a significant part in the response to the HA Test. The presence or absence of conspecifics in a neighbouring pen have been found to have a small influence on responses to handling tests (Grignard *et al.*, 2000), showing that social separation can play a role in animals' responses to tests.

5. Is there a relationship between the FF Score and the time spent in contact with the wooden boards in the SS Test?

If investigatory motivation plays a significant part of the response to the FF Test in the tamer animals, a relationship may be seen between the responses of the animals to the FF Test and how long they spent in contact with the wooden boards in the SS Test, which may also be a measure of investigatory motivation. The wooden boards were introduced to the animals a couple of days before the tests were carried out, so that they would not be novel to the animals in the test. However, they still elicited quite a lot of interest, and some of the animals spent time in the test sniffing, licking or rubbing their head against the boards.

There was no relationship between these two measures in the bulls (Spearman Rank correlation; $n = 41$, $r_s = 0.15$, $p = 0.348$) or the heifers ($n = 41$, $r_s = 0.06$, $p = 0.660$). This comparison therefore provides no evidence that differences in investigatory responses play a significant part in the responses to the FF Test. However, as we don't know that time spent in contact with the wooden boards is definitely indicative of an investigatory motivation, the possibility cannot be ruled out.

4.3. Validation of the Social Separation Test

4.3.1. Methods
4.3.2. Results and Discussion

4.3.1. Methods

Three behavioural tests, the FF Test, SS Test and HA Test, were carried out on the Year 3 heifers born in Spring 2001 as part of the experiment described in Chapter Six. Additionally, another test, the Sociality (SO) Test was carried out on 42 of the heifers. The objective of this experiment was to compare the results of the SO Test with those from the SS Test, to see if there was a strong relationship between the two, and hence validate the SS Test.

At the time of testing, the heifers were housed in six groups with 10 - 13 animals in each group. The six groups were labelled K, L, M, N, O and P in order of approximate average weight (and hence age). The median age of the heifers on the first day of testing was 309 days (range 251 - 338 days).

The four tests were carried out according to the schedule shown in Table 2.14. The procedures for the SS and SO Tests are described in Chapter 2.3.3 and 2.3.6. The FF Tests were carried out first over the course of four days, with between one and three days spent on each group. The SS and SO Tests were then carried out on two groups at a time. Two repeats of the SS Test were carried out on consecutive days, and five days later, SO tests were carried out on three consecutive days. The SO Test was carried out on the first four groups to be tested; K, L, N and P. The first two animals from group K were used in a short pilot trial to check the test procedure and were not included in the experiment.

The median, range and distribution of each test measure were examined for variation between animals. Differences in scores between repeat tests were investigated using Paired t-tests. Repeatability coefficients over repeat tests were estimated as described in Chapter 2.5.2. Spearman Rank correlation coefficients were also calculated over repeated tests, and were also used to compare relationships between measures from the two tests.

4.3.2. Results and Discussion

Although FF, SS, HA and SO Tests were carried out on the Year 3 heifers, only the results of the SS and SO Tests carried out on 42 heifers from groups K, L, N and P are discussed here. When testing one of the animals, the second SS Test was disturbed and was discounted, so the analysis described below was carried out on 41 animals.

The duration of time spent Walking, showing Escape behaviour and Running (WER) by the heifers in the SS Tests was used as a measure of sociality. WER ranged from 9 - 215 seconds with a median of 80, and showed a positively skewed distribution. No change in levels of WER was seen between the two repeats of the test (Paired t-test: $n = 41$, mean difference = -10.3 seconds, $t = -1.43$, $p = 0.159$). Repeatability of WER between the two repeats was calculated, and found to be high, at 0.75 ± 0.07 ($n = 41$, deviance difference = 33.92, $p < 0.001$). This result was backed up by a high correlation between WER from the two repeat tests (Spearman Rank correlation: $n = 41$, $r_s = 0.83$, $p < 0.001$). These results are similar to those found on the 54 Year 1 heifers in Chapter 3.3.2.3, with slightly higher repeatability shown by the Year 3 heifers.

An additional measure was taken from the SS Test for possible comparison with the measures in the SO Test - the total test time spent by the animals in the quarter of the pen closest to their pen-mates ('Buddy Time'). This was measured as it was the closest equivalent measure to the measure of 'Accumulated Time in Close Zone' in the SO Test. Buddy Time ranged from 64 - 281 seconds and was normally distributed with a mean of 171 seconds. No change in the levels of Buddy Time were seen when the measures were compared between the two test repeats (Paired t-test: $n = 41$, mean difference = -12.5, $t = -1.19$, $p = 0.241$). However, this measure was not repeatable between the two test repeats ($n = 41$, $r = 0.15 \pm 0.15$, deviance difference = 0.93, $p > 0.5$). This estimate was backed up by a lack of correlation (Spearman Rank correlation: $n = 41$, $r_s = 0.16$, $p = 0.334$). Hence Buddy Time was not a useful measure, and so was not used for inter-test comparisons.

Three measures were recorded in the SO Test, 'Latency to first chalk line' (an 18 m distance), 'Latency to enter close zone' (a 27 m distance), and 'Accumulated time in close zone'. For a reminder of the distances and layout of the test, see Figure 2.12 in Chapter 2.3.6. As this was a test that had not been previously examined in Chapter Three, it was repeated on each heifer three times, and variability and repeatability of the measures were examined here. All three variables showed a wide range of response (see Table 4.2), and non-normal distributions due to a large number of small values. All of the variables were moderately repeatable (see Table 4.2). 'Latency to first chalk line' ('Latency to Line') and 'Accumulated time in close zone' ('Close Time') showed the highest repeatability values, of 0.38 and 0.41 respectively, and these two variables were selected for further analysis. The analysis statistics are summarised in Table 4.2.

An increase in mean values of Latency to Line was seen between Tests 1 and Test 2 (Paired t-test: $n = 42$, mean difference = -22.3 seconds, $t = -2.48$, $p = 0.017$), and no change in response was seen between Tests 2 and 3 ($n = 42$, mean difference = 10.2 seconds, $t = 1.24$, $p = 0.222$). Animals were slower to move to the area of the corridor closest to their penmates when the test was repeated a second time. Across the three tests, the repeatability of the Latency to reach the first line was moderate at 0.38. When calculated over just the first two tests, the repeatability coefficient was 0.30, and across Tests 2 and 3 was 0.40. Correlations between test repeats also showed that moderate relationships existed between Tests 1 and 2 ($r_s = 0.49$) and Tests 2 and 3 ($r_s = 0.52$). Together, the results show that most consistency of response was seen between the second two repeats of the test. Similar patterns of response were seen in the other tests assessed in Chapter Three. It was concluded there that the novelty of each test situation played a part in the animals' responses to the first of each test. This is even more likely to be the case here, as the corridor where the test was carried out was relatively unfamiliar to the animals. However, the repeatability of Latency to Line was only moderate.

The same analysis was repeated on the variable 'Close Time'. No difference in Close Time was found between successive test repeats (Paired t-test: Tests 1 & 2, $n = 42$, $t = 1.70$, $p = 0.097$; Tests 2 & 3, $n = 42$, $t = -0.98$, $p = 0.333$), showing that the animals didn't spend more or less time in the close zone in successive test repeats. Repeatability of Close Time across the three tests was 0.41. When calculated across the first two tests, the repeatability was high at 0.53, and across Tests 2 and 3 was also high at 0.50. Correlations between test repeats also showed moderate positive relationships between Tests 1 and 2 (r_s = 0.55), and Tests 2 and 3 (r_s = 0.54). Therefore Close Time was a more repeatable measure than Latency to Line, and repeatable enough to be considered a useful trait measure.

Table 4.2: A statistical summary of the measures taken from 42 Year 3 heifers in the Sociality Test.

Statistic		Test Measure		
		Latency to Line	Latency to Close Zone	Close Time
Min (seconds)		5	7	0
Max (seconds)		180	180	173
Median (seconds)		11	26	75
Repeatability	(3 tests)	0.38 ±0.10 (15.76***)	0.25 ±0.10	0.41 ±0.10 (19.33**)
±s.e.	(Tests 1+2)	0.30 ±0.14 (3.85*)	/	0.53 ±0.11 (13.51***)
(deviance diff)	(Tests 2+3)	0.40 ±0.14 (7.08***)	/	0.50 ±0.12 (12.29***)
Spearman rank	(Tests 1+2)	0.49***	/	0.55***
correlation	(Tests 2+3)	0.52***	/	0.54***

*** $p < 0.001$, ** $p<0.01$, * $p<0.05$.

Finally, results from the SS Test were compared with results from the SO Test to determine if there was a relationship between the heifers' responses to the two different situations of separation from the herd (Fig. 4.7 and Fig. 4.8). Correlation coefficients were calculated between WER from the SS Test and Latency to Line and Close Time from the SO Test. The results from the second repeat of each test were used. WER was not correlated with Close Time (r_s = 0.261, p = 0.099) and showed a trend towards a negative correlation with Latency to Line (r_s = -0.298, p = 0.058), indicating that heifers showed a high level of sociality in the SS Test tended towards taking a short amount of time to move towards their penmates in the SO Test. However, the relationship is only small, and Figure 4.7 shows that there is no clear relationship between the two variables. Therefore the SS Test measure has not been validated by comparing it to measures from the SO Test.

It is possible that the SO Test procedure was not suitable for validating the SS Test, because it was carried out in an area of the building that was not as familiar to the animals as their home pen. The novelty of the test corridor may have influenced the animals' responses, even in the second and third repeats of the test. Hence it may have been more appropriate to thoroughly habituate the animals to the corridor, possibly by penning them there for a few hours at a time on the days leading up to the SS Tests. Unfortunately this was not possible at the time.

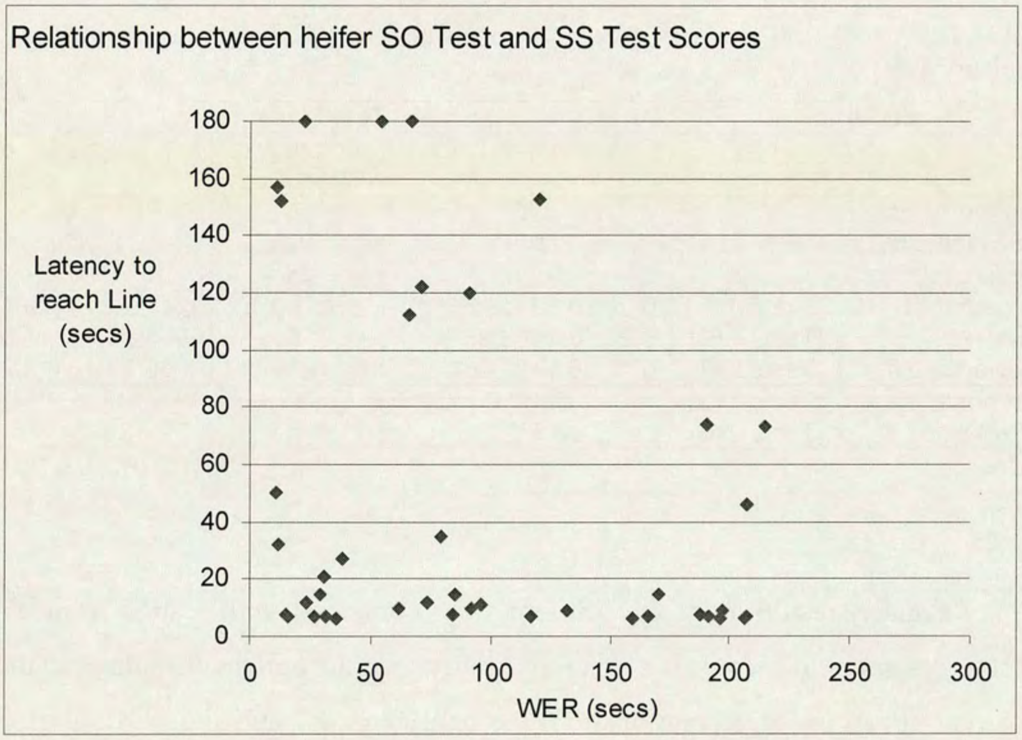


Figure 4.7: Relationship between 'Latency to reach Line' from the second Sociality Test and durations of WER (Walk, Escape and Run) from the second Social Separation Test carried out on 41 Year 3 heifers.

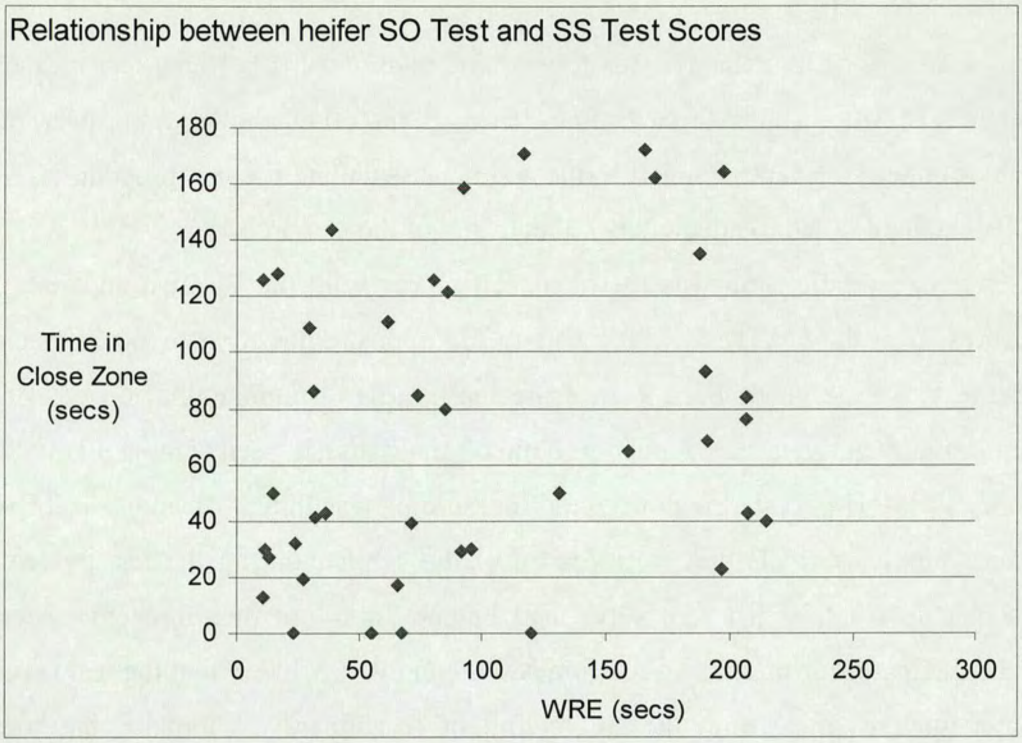


Figure 4.8: Relationship between 'Time in Close Zone' from the second Sociality Test and durations of WER (Walk, Escape and Run) from the second Social Separation Test carried out on 41 Year 3 heifers.

4.4. Conclusions

The aim of this chapter was to examine relationships between test measures from the FF, HA, SS and SO Tests, in order to interpret the behavioural traits seen in terms of underlying temperament traits, and hence validate the test procedures. The results obtained failed to adequately validate any of the test procedures.

A low relationship was seen between scores from the FF Test and Latency measures from the HA Test, which was in the opposite direction to that expected. These tests were both thought to measure fearfulness of humans, but clearly other factors significantly affected one or both of the tests. It seems most likely that Latency in the HA Test was not a good measure of fearfulness of humans, for two reasons. Firstly, animals that were fearful of the handler during the test procedure were able to react in different ways, and Latency may not therefore have been a suitable indicator for measuring fearfulness. Secondly, it is likely that the test is not a good method of measuring the single trait of fearfulness of humans, but rather measures ease of handling of animals due to a mixture of traits, such as fearfulness and aggression. Chapter Three showed Latency in the HA Test to be a highly repeatable measure, in the heifers at least, and therefore a very useful practical measure for discriminating between heifers on the basis of their ease of handling, whatever the underlying cause(s). The FF Test is likely to be the more useful of the two tests in terms of taking measures of single temperament traits, but it requires further validation before we can be sure about its interpretation as a test of fearfulness of humans.

No relationship was seen between measures from the SS Test and the SO Test, which were both thought to be measures of sociality. Again, measures from both tests were highly repeatable, showing that they are both measures of some trait, or mix of traits. Unfortunately, as with the FF and HA measures, it has not been shown that the responses can be regarded as anything more than consistent responses to the very specific test situations imposed on the animals. Interpretations of the test results must be made with care, until further validation of these types of tests can be made.

The SS Test clearly measured a different trait from both the FF and HA Tests. The lack of relationships between the test measures support an assumption made at the start of this project, that the fear responses of cattle to different stimuli, such as humans, novelty and social separation, are stimulus-specific, independent traits. The two points of view concerning whether fearfulness is a uni- or multidimensional trait were described in Chapter 1.3. The results of this study disagree with the 'unity of fear' concept that is supported by some authors of cattle temperament studies (e.g. Boissy, 1995; 1998).

Studies comparing the reactions of cattle in different test situations *have* found relationships between tests that cause fear using apparently different stimuli (Boissy & Bouissou, 1988; 1995). In each of these, the authors offer alternative hypotheses to explain their findings. The first is that the different tests all measure the same underlying trait, that predisposes the individual to respond with the same strength to all potentially challenging events. The alternative is that similarities in the behavioural responses to the different tests are caused by stressful elements that are common to all the tests carried out. I believe that relationships found between different test measures in these studies are best explained by the latter hypothesis.

Boissy and Bouissou (1988) tested the influence of early handling on the behavioural responses of Friesian heifers to a variety of tests. Some of the tests were designed to involve the presence of a human (feeding in an open-field test with a human present, a flight test, a sorting test, a leading test and a tethering test). Others were designed so that they did not (clearing an obstacle, an open-field test, feeding in an unfamiliar environment). Heifers that were handled more than control animals during rearing were less reactive to the tests involving human contact, but, with the exception of one group in one test, they were *not* more reactive to the tests that did not involve human contact. This finding suggests that prior handling habituated the animals to the fearful stimulus of human contact, but did not lessen their reactions to other fearful stimuli, suggesting the responses to the different types of tests were independent.

Boissy and Bouissou (1995) again compared Friesian heifers' reactions in different test situations: a novel environment, a novel object, a conflictual situation, and a surprise affect. Moderate correlations between responses in the different tests were found. However, all the tests included social isolation, and three were based on novelty. The authors stated that further experimentation was needed before concluding that heifers show consistency in a range of fear-inducing situations.

The conclusions of the two experiments described in this Chapter are that, although the FF and SS Tests measured consistent properties of the animals that may yet be shown to be temperament traits, interpretations of the measures must be treated with caution in the meantime. The HA Test, while giving consistent results, is likely to reflect a mix of temperament traits and is unlikely to be useful for the measurement of a single trait. The FF Test is likely to reflect fearfulness of humans, but requires further validation. The tests in this study were carefully designed to minimise (as much as possible) stimuli other than the one of interest affecting the animals in each test situation. The results obtained lend further support to the hypothesis that fearfulness is a multidimensional trait in cattle.

Chapter Five: Temperament Traits in Cattle over Time

5.1. Introduction

One of the most interesting applications of the study of temperament is the potential use of measures taken at an early age to predict the future manageability of animals. Results from behavioural tests carried out on animals at a young age may have the potential to predict levels of ease or difficulty of handling, due to low or high levels of fearfulness, that will persist in the animals when older. One useful application of this could be the prediction of ease of handling and milking in the dairy parlour. Dairy cows are culled on the basis of their unsuitable behaviour in the dairy parlour. Kicking and reluctance to enter the parlour or stand still slows down the milking process, disturbs other animals, and presents risk of injury to the stockpersons. Temperament traits such as fearfulness of humans, neophobia and aggression are likely to contribute to such behaviour. If it were possible to use a behavioural test to identify individuals at an early age, that had a high chance of being difficult to handle for milking later on, they could be removed from the herd much earlier. Raising replacement heifers for dairy herds is very expensive, and so removing problem animals at an early age could result in significant savings.

As previously noted (Chapter Three), few studies of these temperament traits in cattle using behavioural tests have looked at the repeatability of animals' behavioural reactions when tested more than once. Even fewer have looked at consistency of responses over long periods of time. Fisher *et al.* (2000) carried out the only study on the repeatability of flight tests over long periods of time, carrying out three test repeats at one month intervals. Repeatability values obtained were moderate to high. A handling test was carried out on a group of heifers at nine months of age, and then again seven months later (Grignard *et al.*, 2001). A substantial relationship was found between the Docility Scores from the two tests. Tests involving separating individual cattle from their conspecifics have not been repeated on the same animals using intervals longer than consecutive days (Hopster & Blokhuis, 1993). Although Le Neindre (1989) carried out open-field tests on cows a year apart, the study concentrated on comparing group differences, and made no reference to correlations between the test results. This chapter therefore addresses whether the traits measured using the tests in this thesis are consistent over long periods of the animals' lives. There are two aims to the chapter. The first was to investigate whether the behavioural tests used in this thesis gave consistent results when carried out on the same animals at different ages. The objective was to compare behaviour displayed by a group of 49 heifers in the Social Separation (SS) Test and the Handling (HA) Test when tested at approximately four months of age, and then again at approximately 12 months of age.

The second aim of this chapter was to investigate whether responses obtained from the behavioural tests measuring fearfulness of humans are predictive of future behaviour while being milked in the dairy parlour. In the last few years, the behaviour of dairy cattle while being milked has come under increasing attention, due to the links found between behaviour of stockpersons towards their cows, subsequent behaviour of the cows and milk productivity measures (Rushen *et al.*, 1999; Breuer *et al.*, 2000; Hemsworth *et al.*, 2000; 2002; Waiblinger *et al.*, 2002). Behavioural responses to milking vary between animals, and are thought to be associated with fearfulness. Willis (1983) reported that during milking, some cows constantly flinch their udders or stomach muscles, shift their weight from one hind foot to the other and kick (the 'FSK response'). While this behavioural complex is probably a response to a number of stimuli involved in the milking process, it is likely that fear of humans is a major component. This is supported by the finding that handling by humans around the time of calving reduced the number of FSK responses that primiparous cows displayed during the first two weeks of milking (Hemsworth *et al.*, 1987; 1989). A relationship has been found between FSK responses in the dairy parlour and flight distances (Purcell *et al.*, 1988) but this association was not confirmed in later studies (Breuer *et al.*, 2000; Waiblinger *et al.*, 2002). The objective of this study was to determine whether the behaviour of 32 heifers scored in the Flight-from-Feeder (FF) and HA Test at 11 months showed a relationship with FF Test Scores and measurements taken in the dairy parlour in the first few weeks after calving, at approximately 30 months of age.

5.2. Stability of Temperament Measures between four and 12 months

5.2.1. Methods

A series of three behavioural tests were carried out on heifers born in Year 2 of the project, when they were approximately four months of age. A second series of the tests was carried out on the same heifers at approximately 12 months of age. The analysis presented here is based on data from 49 animals which were tested at both ages. The median age of the heifers when the first series of tests began was 119 days (approximately four months), with a range of 98 - 138 days. The median age of the heifers when the second series of tests began was 380 days (approximately 12 months), with a range of 328 - 409 days. The test procedures used are described in Chapter 2.3, and the testing schedules followed are described in Chapter 2.4. Briefly, three SS Tests, one Novel Object (NO) Test and three HA Tests were carried out at four months, and two FF, two SS Tests, two NO Tests and two HA Tests were carried out at 12 months. The repeat tests were carried out on consecutive days. Of the tests that were repeated three times, only the first two test repeats of each were analysed. This decision was taken as the SS and HA Test were previously shown to give highly repeatable results when carried out more than once on the same animals, and the measures taken from the second test repeats were thought to be the most useful (see Chapter 3.5).

The results from the NO Tests carried out on the Year 2 heifers at 12 months have already been discussed in Chapter 3.4.3. The results showed little variation and poor repeatability, showing that this test did not provide useful measures of temperament traits in these animals. Therefore, it was decided that it would not be useful to make comparisons with the data from the animals at four months. The NO Test is not discussed further in this chapter. FF Tests could not be carried out on the heifers at four months as they were not big enough to use the automatic feeders at this age, so FF Tests are also not compared across the two ages.

The main measure taken in the SS Test was the duration of time spent Walking, showing Escape behaviours or Running (WER). This measure was chosen after Principal Components Analysis (PCA) had suggested that these behaviours were indicative of sociality in the Year 1 animals, with a high duration implying a high sociality motivation (see Chapter 3.3.2.1). The WER Scores obtained were compared with scores from another test thought to measure sociality carried out on these animals (see Chapter 4.3). The lack of a relationship found between the two tests means that the SS Test has not been validated. The hypothesis that the test measured sociality remains, and is interpreted as such in this experiment, but this interpretation must be treated cautiously.

In the present study, PCA was carried out on the durations of all the behaviours displayed by the Year 2 heifers in the SS Tests at both ages. The eigenvectors (loadings) of the two main components from each analysis were plotted, in order to produce a graphical presentation of clusters of behaviours that may share a common motivational background (see 2.4.4). This allowed the investigation of two issues. Firstly, whether the same pattern of clusters of behaviours (especially W, R and E) were seen in the Year 2 heifers at 12 months as was seen in the Year 1 heifers of the same age, described in Chapter 3.2. Secondly, whether the same pattern was seen in the Year 2 heifers at both the younger and older age. The medians and ranges of WER, comparisons between first and second test values, and repeatability estimates were then compared between the two ages. Finally, Spearman Rank correlation coefficients were calculated in order to investigate whether a relationship could be found between the WER Scores from the second repeat of each test carried out at four months, and the second of each test carried out eight months later. The second repeat of each test was considered the most useful for these comparisons, because the novelty of the test situations seems to impact on the responses to the first test carried out each time (see Chapter 3.5).

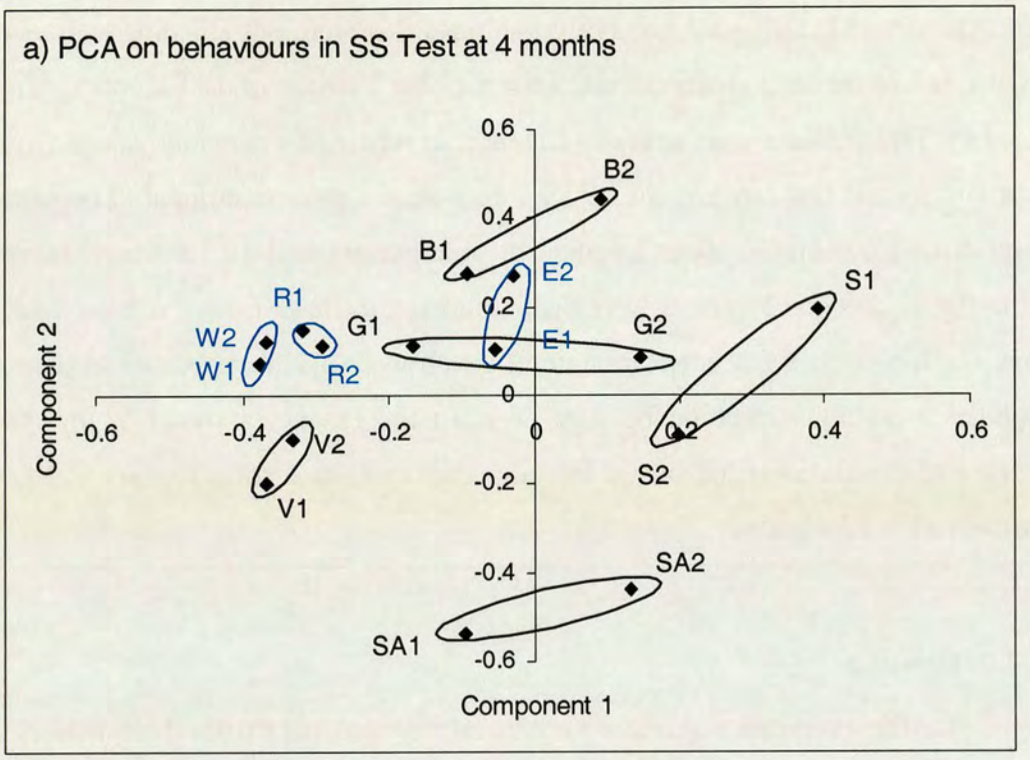
The HA Test measure of 'latency to hold the animal in the test corner for 30 seconds' ('Latency') gave a measure of how easy each animal was to handle, and was probably a measure of a mix of underlying traits such as fearfulness of humans and aggression (see Chapter 4.2.2). The test lasted for two minutes, and an extra 30 seconds was added onto the Latency of animals that were not standing in the corner by the end of the test, giving a maximum possible Latency of 150 seconds (Chapter 3.3.4.1). The medians and ranges of Latency, comparisons between values from the first and second test repeats, and repeatability estimates were compared between the tests carried out on the Year 2 heifers at four months and 12 months. The results from the heifers at 12 months were also compared to those from the Year 1 heifers. Finally, Spearman Rank correlation coefficients were again calculated to determine whether a relationship could be seen between the Latencies shown in the second repeats of the tests carried out at four months, and the second repeats of the tests carried out at 12 months.

5.2.2. Results

Of the 49 animals tested at four months of age, data from three heifers were not available from the first SS Test, and data from two of these were not available from the second SS Test, due to tests being disrupted. Data from the tests at 12 months were not available from one animal from each of the first and second tests, as they were showing behavioural signs of oestrus at the time of testing.

5.2.2.1. PCA of behaviours in SS Test

Behaviours displayed in the SS Test were recorded as detailed in the ethogram in Table 2.3. The durations of eight behaviours states (Escape (E), Gambol (G), Kneel (K), Run (R), Stand alert (SA), Stand occupied (S), Stand and sniff/lick/rub boards (B), Walk (W)), and the frequency of Vocalise (V) were analysed using PCA. Plots of the eigenvectors (loadings) of the first and second components from the heifer data at the two ages are shown in Figure 5.1. All the animals that were scored in both test repeats were included, which gave 46 animals at four months and 47 animals at 12 months. The plots show two loadings for each behaviour circled, representing the values from the two repeats of the test. At four months the first and second components explain 26 % and 13 % of the total variation in the data respectively, and at 12 months, similar proportions of 27 % and 14 % respectively.



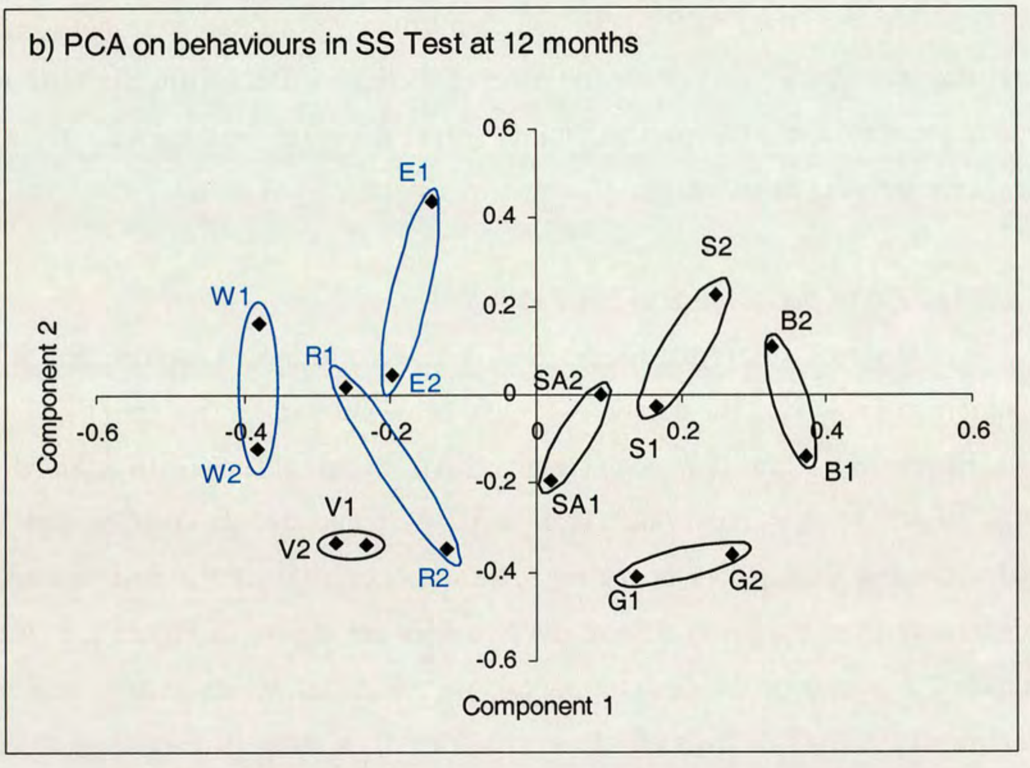


Figure 5.1: Graphs showing plots of the eigenvectors of the first and second components from Principal Components Analysis carried out on behaviours shown in SS Tests on the Year 2 heifers, a) at 4 months (n = 46), and b) at 12 months (n = 47). A value for each behaviour from each of the two repeats of the test is shown (and circled). E = escape, G = gambol, R = run, SA = stand alert, S = stand occupied, B = stand and sniff/lick/rub boards, W = walk, V = vocalise. The total duration of W, R and E was the chosen test measure, and these behaviours are circled in blue.

The PCA loadings of the Year 1 heifers at 12 months and the Year 2 heifers at 12 months show a similar pattern when compared (see Figures 3.3b and 5.1b). W, R and E cluster together in each graph, and load strongly on Component 1, implying that they contribute a large part of the variation represented by this component. To investigate this in the Year 2 heifers, a Spearman Rank correlation was carried out between the WER Scores from Test 1 and Test 2, and the Component 1 scores from a PCA carried out on Tests 1 and 2, respectively. Very high correlations were found ($n = 47$; SS Test 1, $r_s = -0.964$, $p < 0.001$; SS Test 2, $r_s = -0.949$, $p < 0.001$). Therefore Component 1 can be hypothesized as representing 'sociality'. The justification for selecting the combined duration of WER as a measure of sociality in the Year 1 heifers was described in Chapter 3.3.2. Figure 5.1b shows that WER is also a valid measure of sociality in the Year 2 heifers, at least at this age.

The heifers also showed a similar pattern when they were tested at four months old (Figure 5.1a). W, R and V lie at one end of Component 1, contrasting with S, which lies at the other. However, E does not fall into a cluster with W and R as it does when the heifers are older. Closer examination of the data revealed that the heifers spend very little of the test time displaying Escape behaviours at this age (median 3.2 seconds, range 0 - 39 seconds) compared to when they were tested eight months later (median 29 seconds, range 0 - 94 seconds). The levels of E are so low at four months that the behaviour does not contribute much to the total variation in the data, and so does not load strongly on either component in the PCA. Nevertheless, as E showed the same direction of loading as W and R in both components, it was decided to use the WER measure at this age also, and this enabled direct comparisons with the data from when the heifers were tested at an older age.

5.2.2.2. Analysis of WER from the SS Test

The median and range of the duration of WER from the first and second repeats of the SS Tests were calculated at both ages. At four months the observations were normally distributed, and at 12 months they showed a slight negative skew. The medians and ranges shown by the Year 2 heifers in Tests 1 (median 175.4, range 8.3 - 244.3) and 2 (median 183.7, range 16.8 - 262.8) at 12 months were very similar to the heifers tested the year before (see Chapter 3.3.2.3). These levels are much higher than the animals showed at four months of age, both in the first test repeat (median 81.9, range 5.5 - 168.6, Paired t-test; $n = 45$, mean difference = 81.8 seconds, $t = -8.12$, $p < 0.001$), and the second test repeat (median 105.5, range 17.4 - 201.2, Paired t-test; $n = 46$, mean difference = 65.0 seconds, $t = -7.15$, $p < 0.001$). This indicates that the heifers exhibit lower levels of sociality at the younger age. An increase in WER Score was seen when the heifers were tested a second time on the following day at four months of age (Paired t-test; $n = 46$, mean difference = -18.1, $t = -3.74$, $p < 0.01$) but not when the test was repeated a second time at 12 months ($n = 47$, mean difference = 2.2 seconds, $t = 0.25$, NS). Therefore levels of sociality were higher on repeat testing at the younger age, but not the older.

Repeatability values were calculated between the first and second tests at both ages. The WER Scores taken from the Year 2 heifers showed high repeatability estimates, both at four months of age ($r = 0.65 \pm 0.09$, $n = 46$, deviance difference = 24.15, $p < 0.001$) and at 12 months of age ($r = 0.53 \pm 0.11$, $n = 47$, deviance difference = 14.96, $p < 0.001$). These values demonstrate that the test provides useful information about an inherent trait of the animals at both ages.

Finally, the scores from the second tests carried out at the young and older ages were compared to look for relationships between the two testing periods (see Fig. 5.2). WER Scores from the second SS Test carried out at four months of age showed a moderate correlation with scores in the second SS Test eight months later (Spearman Rank correlation, $n = 46$, $r_s = 0.39$, $p = 0.007$).

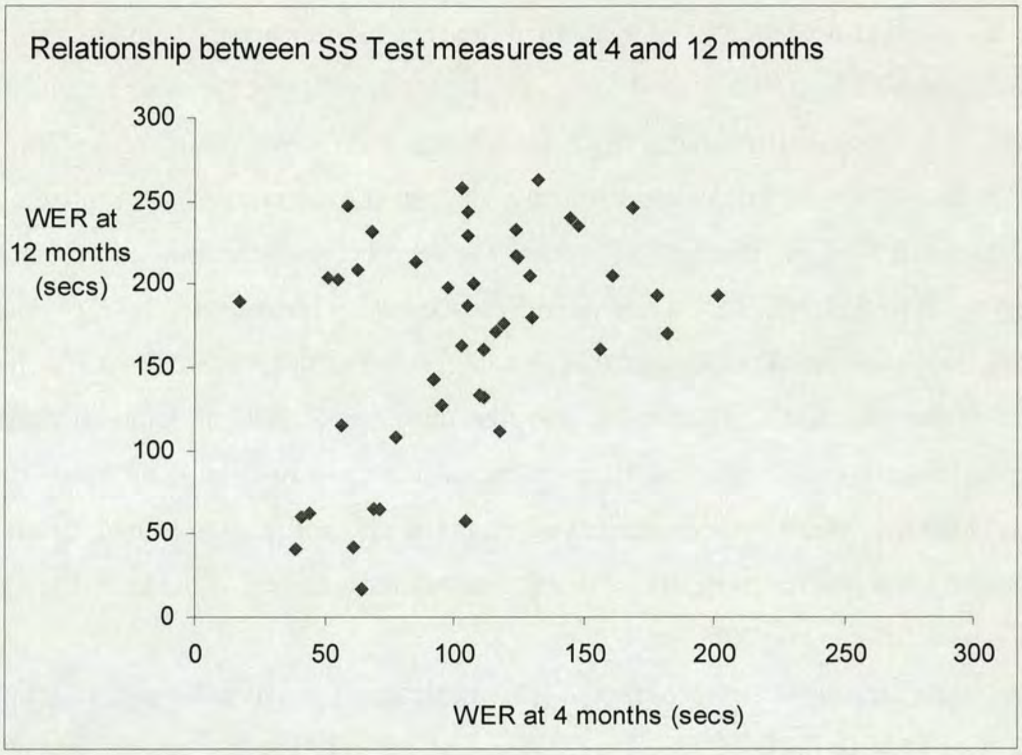


Figure 5.2: Graph showing the relationship between SS Test measures from the second repeat of each test carried out on 46 Year 2 heifers at 4 months, and at 12 months. WER = total duration of Walking, Running and Escape Behaviours shown during the test.

5.2.2.3. Handling Test

HA Tests were carried out on all 49 of the heifers, twice at each age. The median and range of Latency from both the first and second repeats of the HA Tests were calculated at both ages. The Latencies were not normally distributed as the scale used was limited. The median Latency shown by the heifers at 12 months was 113.3 in Test 1 (range 30.3 - 150.0), and 51.9 in Test 2 (range 12.1 - 150.0). These values are similar to those shown by the Year 1 heifers (see Chapter 3.3.2.3). A decrease in Latency is seen between the first and second tests (Paired t-test; $n = 49$, mean difference = 35.3 seconds, $t = 5.03$, $p < 0.001$), implying that the heifers are less fearful during the second test. Repeatability values were calculated between the first and second tests. The Latencies showed moderate repeatability values at 12 months ($r = 0.34 \pm 0.13$, $n = 49$, deviance difference = 5.88, $p < 0.05$).

These Latencies are longer than those that were shown by the same heifers when tested at four months of age, when the first test repeats are compared (median 68.5, range 10.5 - 150.0, Paired t-test; $n = 49$, mean difference = -30.2 seconds, $t = -2.93$, $p < 0.01$) and the same when the second test repeats are compared (median 47.5, range 7.3 - 150.0, Paired t-test; $n = 49$, mean difference = -6.5 seconds, $t = -0.65$, $p < 0.516$). No change in Latency was seen between the first and second test repeats at four months ($n = 49$, mean difference = 11.6 seconds, $t = 1.47$, $p = 0.147$). Together, these results suggest that the heifers were the most difficult to handle during the first test at 12 months, and that during both tests at 4 months, and the repeated test at 12 months, they were easier to move into the test corner. Repeatability values were calculated between the first and second tests. Latency also showed moderate repeatability values at four months of age ($r = 0.38 \pm 0.12$, $n = 49$, deviance difference = 7.68, $p < 0.01$).

Finally, correlation coefficients were calculated between the tests results from the two ages (see Fig. 5.3). Latency from the second HA Test carried out at four months of age showed no relationship with Latency Scores in the second HA Test eight months later (Spearman Rank correlation, $r_s = 0.08$, $n = 49$, $p = 0.569$).

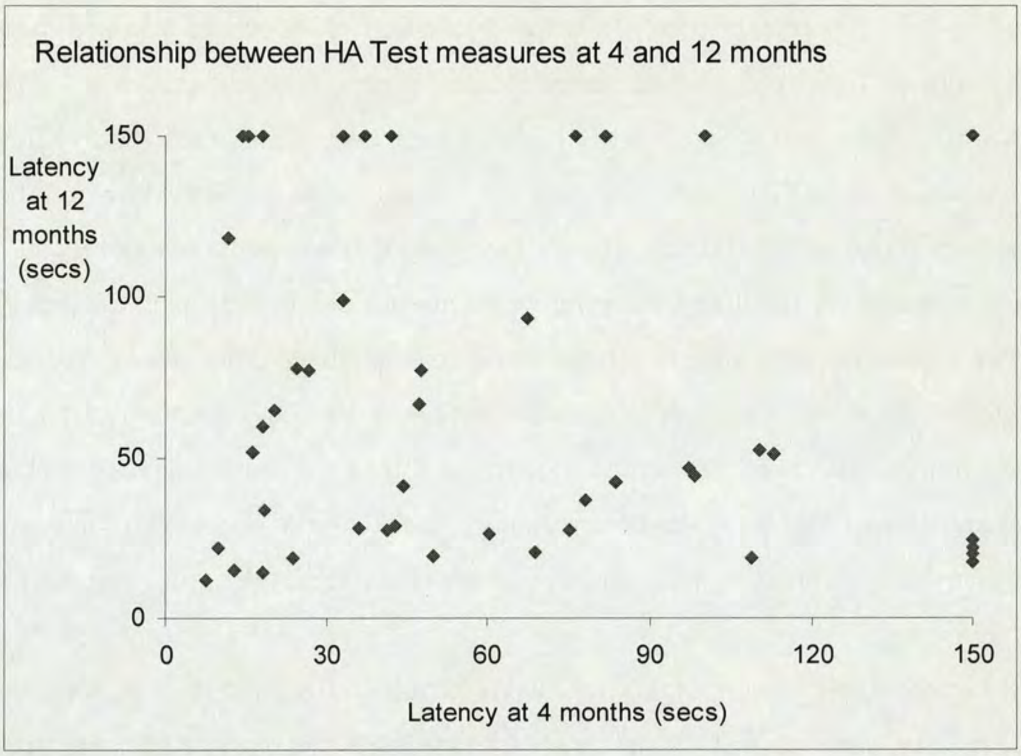


Figure 5.3: Graph showing the relationship between HA Test measures from the second repeat of each test carried out on 49 Year 2 heifers at 4 months, and at 12 months. Latency = time taken for the animal to stand in the corner for 30 seconds.

5.2.3. Discussion

This experiment examined whether responses seen in two behavioural tests were consistent when the tests were carried out at two ages, eight months apart. It was assumed that carrying out the tests at four months would not affect the responses shown when the tests were repeated at 12 months. The animals had no contact with the person who handled them in the HA Tests in the intervening period. Results from the tests at 12 months were compared with the results from the Year 1 heifers who were tested for the first time at that age, and similar WER levels and HA Latencies were seen in both years.

Levels of WER seen in the SS Test are thought to indicate levels of sociality. The heifers showed much lower levels of these behaviours at four months than at twelve months. It is not clear why such a difference was seen. It is possible that that calves showed lower levels of sociality at the younger age because they were less bonded to the other members of the group than when older. When a mixed-breed suckler herd that enjoyed minimal interference by man was studied, calves were found to spend much of their early life lying on their own in sheltered spots (Kiley-Worthington & de la Plain, 1983). Calves did not initiate many contacts with other herd members until about three months of age, when they showed a marked increase in social contacts. By the time they were twelve months old, the calves in the present study were more likely to have developed close associations to other calves. It is also possible that the calves were more inquisitive at four months, and this motivation had a higher impact upon behaviour seen in the test at this age. The calves in the suckler herd study showed highest levels of investigatory behaviour at one week of age, and levels then declined to six months, when the study ended (Kiley-Worthington & de la Plain, 1983).

Levels of WER were highly repeatable across consecutive days at both four and 12 months, showing that it is a reliable trait measure of the animals at both ages. A substantial relationship was also seen between the measures at the two ages. However, the relationship was moderate, rather than high, as seen when repeatability was measured on consecutive days (reminder; a repeatability estimate of 0.5 is equivalent to a correlation coefficient of approximately 0.7). This implies that the heifers' responses to the test did change slightly between the two ages, and the changes were not consistent across all the heifers. It is possible that the heifers changed in their response during this time as a result of their experience of various management and handling practices. Because the animals were part of an experimental herd, they were moved out of the pens and handled regularly during this time for weighing, blood sampling as part of another experiment, and other management procedures. However, all the heifers were housed in similar groups within the same shed, were fed and handled by the same stockpersons, and underwent the same procedures, and so these experiences should have affected all the animals in a consistent manner. It is also possible that the test procedures may have been interpreted differently by the animals at the different ages, for example, as mentioned above, the tests may have invoked higher investigatory and lower sociality responses at one age than the other. This example would mean that the test may have measured sociality both times, but with an additional motivation (to investigate) that may have been slightly stronger at one age than the other. These two traits are presumably unrelated, and the animals may have experienced different levels of motivation to investigate, leading to changes in the behavioural responses to the test that would not be consistent across the herd. The same affect may occur if, as another example, levels of fearfulness of humans increased or decreased between the two ages. Two people were present during the SS Tests in order to move the animals between pens, and if their presence caused higher fearfulness at one age than the other, this could again account for changes in responses that would be inconsistent across the herd.

In spite of these possibilities, the WER measure nevertheless showed a substantial relationship between four and 12 months, and therefore reflects a temperament trait that is persistent in calves for at least eight months, and most likely on into adulthood. It is a strong enough relationship that it would be possible to use this test to identify animals at an early age that show extreme levels of this trait, thought to be sociality, and are likely to be difficult to manage when separated out from the group when full-grown.

The Latencies seen in the HA Tests gave a measure of how easy the animals were to handle. The mean Latencies across all the animals were the same during each test repeat at each age, except during the first test at 12 months, when the mean Latency was longer. In Chapter 4.2 it was seen that the trait(s) that Latency measures are unclear. So although it can be concluded that the animals were harder to handle during the first test repeat at 12 months than during the other three tests, it is difficult to deduce why this is. These results contrast with those of Grignard *et al.* (2001) who carried out a handling test on Limousin heifers at nine months and 16 months, and found that the animal got more docile as they got older. The differences between the two studies are likely to be explained by differences in husbandry; the heifers in this study were bucket-reared from 24 hours old, and so were used to close human contact from an early age. The Limousin heifers had been raised at different farms until the age of eight months, and had only been in a free-stables system for one month before their first test (Grignard *et al.,* 2001). The difference in age may also have an effect; the animals in the present study were still very young when the first tests were carried out, compared to the Limousin animals.

Estimates of repeatability were moderate at both ages. However, no relationship was seen between the Latencies measures at the two different ages. This result seems fairly consistent with the pattern shown by the SS Test results. The SS Tests showed high repeatability at each age, but only a moderate relationship across the two ages. As the HA Tests showed only a moderate repeatability at each age, there is not enough consistency of response in the measure for a relationship to be seen between four and 12 months. The lack of repeatability of the Latency measure was discussed in Chapter 4.2.2, and is probably due to the mixture of different traits that this test may be measuring. Again, this result contrasts with that of Grignard *et al.* (2001), who found that measures taken on Limousin animals at nine and 16 months showed a substantial relationship. Along with the differences between the herds mentioned above, the difference in repeatability of the measure may also be due to the different measures used in the two studies. Grignard *et al.* (2001) used a Docility Score, which summarised many of the behaviours displayed during the tests (Le Neindre *et al.*, 1995), whereas the single measure of 'Latency to stand in the test corner' was used in this study. The reasons for this choice were explained in Chapter 3.3.4. The Latency measure therefore does not appear to reflect a temperament trait that is consistent across long periods of time.

5.3. Prediction of dairy parlour behaviour from tests at 11 months

5.3.1. Methods

The aim of this experiment was to investigate whether data from behavioural tests that are thought to measure fearfulness of humans can predict future behaviour in the dairy parlour. The heifers born in Year 1 underwent testing in four behaviour tests at the age of 11 months. Nineteen months later, thirty-two of the heifers were FF tested again, and observed in the dairy parlour during their first few weeks of milking. The behaviour they displayed during the milking process was compared to the scores from the FF Tests carried out at the same time, and the FF and HA Tests previously carried out at 11 months.

The testing schedule at 11 months included the FF, NO, HA and SS Tests, and is described in Chapter 2.4. The results of these tests are described in detail in Chapters Three and Four. The FF and HA Tests only are of interest in this experiment, as they are both thought to relate to fearfulness of humans, a trait which will influence behaviour in the dairy parlour. Briefly, the FF Test is thought to measure fearfulness of humans, and each animal was given a score from a categorical scale of 0 - 6 (see Chapter 2.3.2). The HA Test measure of 'latency of animal to stand in the test corner for 30 seconds' ('Latency') gave a measure of how easy each animal was to handle, probably due to a mix of underlying traits such as fearfulness of humans and aggression (and was described in Section 5.2.1). As also mentioned, the novelty of the test situation was thought to have an effect on the animals' responses the first time each test was carried out (Chapter 3.5). It was therefore decided that it was most appropriate to use the animals' measures from the second repeat of each test, and so the data from the second repeat of the FF and HA Tests were used in the comparisons in this section. As the data from the Year 1 heifers from both these tests were discussed in depth in Chapters Three and Four, only a summary of the median and range of the tests results from the 32 heifers that were also observed in the dairy parlour a year and a half later is presented here.

The heifers were managed as a dairy herd, and calved at approximately 30 months of age, in Autumn - Winter 2001 - 2002. Their calves were removed at 24 hours, and the heifers then entered the milking herd. They were housed in a cubicle pen and fed silage from automatic feeders. The first heifers calved in October 2001, and the rest entered the herd in following weeks as they calved. From when the first heifer calved and entered the herd, recording took place weekly for 17 weeks, excluding one week over Christmas. If the recording day fell on an animal's first day of being milked, she was not recorded until the following week, to ensure that each animal had been through the milking process at least twice before being observed. Hence the first observation on each animal was taken between one and seven days after calving. The objective of the testing was to obtain two FF Scores and three SK Scores from each animal.

FF Tests were carried out once a week between 14:30 and 15:30, and scores were recorded from all the heifers in the milking herd every week until Week 8. Then the heifers were split into two groups, which were rotated between one pen with automatic feeders (necessary for the FF Test) and one pen without, on a fortnightly basis. Hence from Week 9 - Week 17, only half of the heifers were tested each week. By the end of 17 weeks, 31 heifers had been tested in the FF Test at least twice. Approximately one hour after the FF Tests were carried out, the heifers were observed during the afternoon milking, between 16:00 and 17:30 every week. By the end of 17 weeks, 32 animals had been recorded three times in the dairy.

The animals' behaviour in the parlour was scored as described in Chapter 2.3.7. Each animal was observed during three periods of close human contact during the milking process. The number of steps and kicks made by each animal during udder wiping, cluster attachment, and teat dipping was counted and recorded by the observer. These numbers were totalled to give a 'SK' Score for each heifer for each week. Flinches (as described by Willis, 1983) were not recorded as they were difficult to see. Milking was carried out by four stockpersons, up to three of whom were present on any one observation day. The identity of the stockperson carrying out each of the three procedures on each animal was also recorded.

While under observation, seven heifers (in different weeks) required extra attention during the second observed period, cluster attachment, to prevent them kicking the clusters off repeatedly. Depending on the stockperson, the animal's tail was held for a few minutes, or a kick-bar was fitted to prevent the heifer kicking with her hind legs. Both of these measures reduced the SK Score that would have been obtained for that week for the cluster attachment and teat-dipping periods without the extra attention. Hence the SK Scores for these seven heifers were adjusted. Unaffected scores for the first of the three periods, udder wiping, were observed on all animals. An assumption was made that heifers who showed a high SK count during udder wiping would also show high SK counts during cluster attachment and teat dipping. The udder wiping SK counts of each heifer was therefore used to estimate the total unaffected SK count for each, in the following way.

A 'Deviation Score' was calculated for each heifer, which gave an estimate of how much the SK count for udder wiping for that heifer differed from the group mean for the count. It was calculated using the equation:

$$\textit{Heifer Deviation Score} = \frac{\textit{Heifer Udder SK - Cohort Mean Udder SK}}{\textit{Cohort St Dev Udder SK}}$$

where 'Heifer Udder SK' referred to the SK count observed on the heifer during udder wiping, 'Cohort Mean Udder SK' refers to the mean value of the udder wiping SK count observed from all heifers *not* requiring extra attention, and 'Cohort St Dev Udder SK' refers to the standard deviation of the udder wiping SK count observed from all the heifers not requiring attention.

The Deviation Score for each animal was used to calculate the missing Total SK Count for each, that had the same deviation from the group mean as the Udder SK count showed. The following formula was used:

$$\textit{Heifer Total SK} = (\textit{Cohort St Dev Total SK} \times \textit{Heifer Deviation Score}) + \textit{Cohort Mean Total SK}$$

where 'Heifer Total SK' referred to the estimate of the Total SK count that would have been observed with no extra assistance, 'Cohort St Dev Total SK' referred to the standard deviation of the Total SK Count from all the heifers not requiring extra attention, Heifer Deviation Score, which was calculated in the first equation, estimated how much the SK count for udder wiping for that heifer differed from the group mean for the count, and 'Cohort Mean Total SK' referred to the mean total SK Score of all the heifers not requiring extra attention.

These estimated Total SK Scores were used in place of the observed Total SK Scores for the seven heifers. The distribution and range of all the FF and SK Scores obtained were examined. Repeatability coefficients were calculated, and differences between mean scores for the first and second observations were examined using Paired t-tests. Spearman Rank correlation coefficients were calculated to examine whether relationships were seen between measures from the FF Tests at 11 and 30 months, the SK Scores and the FF Scores from 30 months, and finally, the SK Scores and the FF and HA Test Scores from 11 months.

5.3.2. Results

5.3.2.1. FF and HA Test at 11 months

Inter-animal variation and intra-animal repeatability of behaviours shown in the FF and HA Tests carried out on the 56 Year 1 heifers at 11 months were described in detail in Chapter Three, and interpretation of what the tests were measuring was discussed in Chapter Four. Of the 32 heifers included in this study, 29 were tested twice in the FF Test at 11 months, and all 32 were tested twice in the HA Test. The median FF Score was 5, with a range of 1 - 6. The median HA Latency was 50.9 seconds, with a range of 10.5 - 150 seconds.

5.3.2.2. FF Test at 30 months

After calving at 30 months of age, FF Test Scores and dairy parlour behaviour were recorded. The mean age of the heifers on the day of their first observation was 31.8 months (range 29.3 - 34.4 months). Twenty-four of the heifers had two FF Scores from consecutive weeks, and 7 heifers had two scores from between two and five weeks apart (due to the splitting of the group into two housing groups in Week 8, and in some weeks a few heifers were not tested if they did not approach the automatic feeders during the time that testing was being carried out). The median FF Score shown was 6 (range 1 - 6), demonstrating that a large number of the heifers (20) scored 6 in both repeats of the test (see Figure 5.4). A moderate repeatability estimate of 0.45 ± 0.14 ($n = 31$, deviance difference = 6.96, $p < 0.01$) was obtained, indicating that the heifers showed consistency in their responses across the two tests. The scores obtained in the first FF Test were compared to the scores in the second test, and no difference was seen (Paired t-test; $n = 32$, $t = -1.49$, $p = 0.148$).

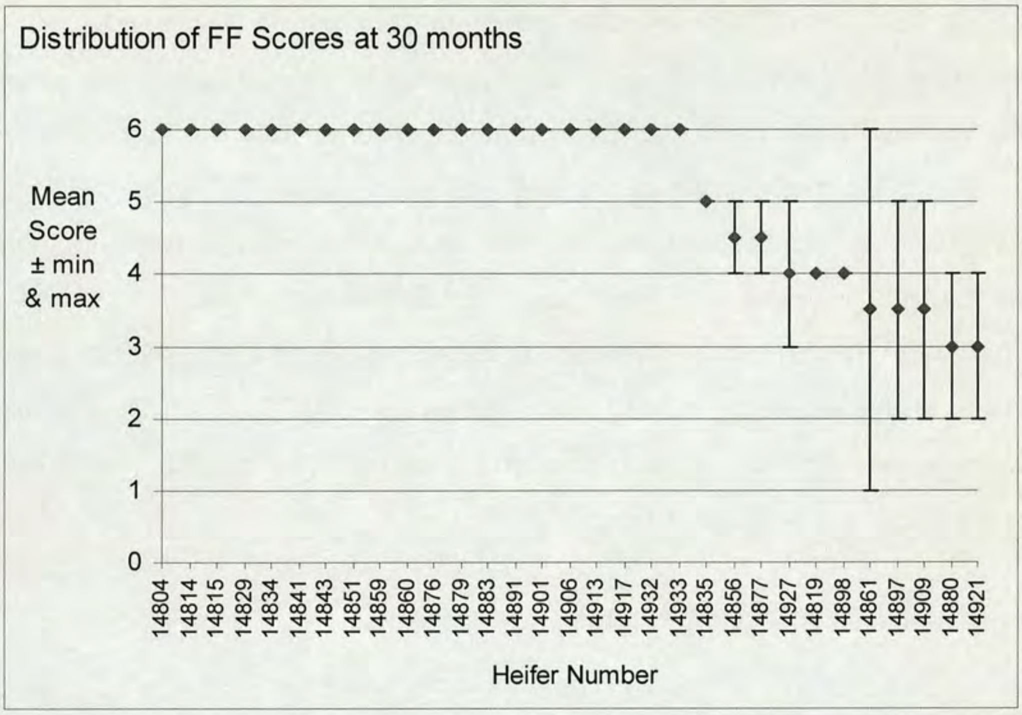


Figure 5.4: Graph showing the mean of two Flight-from-Feeder Test Scores shown by 31 Year 1 heifers at 30 months. The categories range from 1 (animal moved away when observer was within 2.00 - 1.25m) to 6 (animal did not move back when touched). See Table 2.2 for the full scale. The animals are listed in order of decreasing score.

5.3.2.2. SK Scores

Dairy parlour behaviour data from Weeks 3 and 7 were disregarded, as the procedure for udder-wiping was carried out differently during these weeks, meaning that the scores could not be compared with scores from other weeks. If data was missing from an individual heifer from any of the three observed periods on a particular week, the data from that week were also disregarded for that animal. These two reasons, along with the missed week over Christmas, meant that 13 of the heifers had SK Scores from three consecutive weeks, and 19 had SK Scores from three out of four or five consecutive weeks.

The mean score of the SK Scores from each of the heifers from the first three observations were calculated. A wide range of mean SK Scores (0.3 - 25.3) is seen between the heifers (see Figure 5.5). The distribution of the scores is slightly positively skewed. The minimum and maximum bars plotted on Figure 5.5 show that a large number of the animals' scores do not overlap with other animals, demonstrating that each heifer's score seems to be consistent across repeat observations. A repeatability coefficient across all three tests was calculated for the 32 heifers, and a high estimate of 0.58 ± 0.09 (deviance difference = 30.14, $p < 0.001$) was obtained, confirming that the heifers showed consistent inter-animal differences. The mean scores obtained in the first observations of the heifers were compared to the mean scores obtained in the second observations, and no difference was seen (Paired t-test; $n = 32$, mean difference = 0.53, $t = 0.51$, $p = 0.612$). Similarly, no difference was seen between the second and third observations ($n = 32$, mean difference = 0.13, $t = 0.10$, $p = 0.919$).

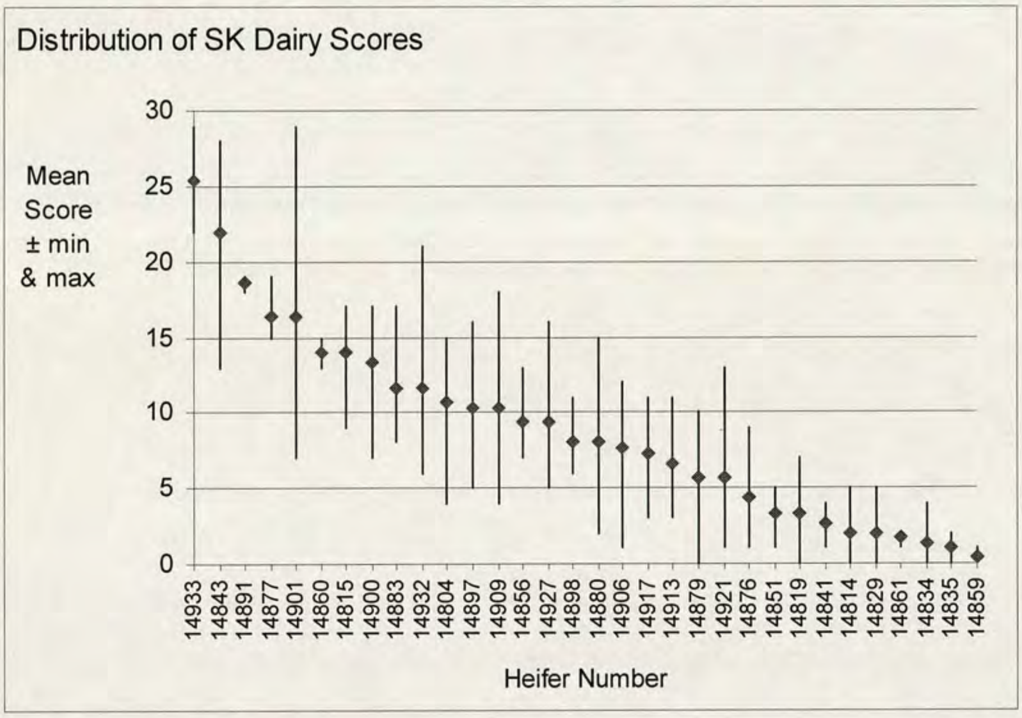


Figure 5.5: Graph showing the mean of three SK Scores observed from 32 Year 1 heifers, at 30 months of age. SK = the total number of steps and kicks counted during three periods of the milking process; udder wiping, cluster attachment and teat-dipping. The animals are listed in order of decreasing score.

5.3.2.3. Relationships between test scores

Finally, relationships between the different scores at the different ages were investigated, using Spearman Rank correlations coefficients. The results from the second repeats of the FF and HA tests were used, as analysis in Chapter Three suggested that the novelty of the test situations contributed to the animals' responses in the first tests. The SK Score from the second observations were therefore also used, to ensure that scores from the first week each heifer was milked were not used, in case the milking parlour was still a novel experience to them at this time (although the analysis above showed that there were no differences between the mean scores from observations 1, 2 and 3). A very high correlation was found between FF Scores from 11 months and 30 months (Fig. 5.6; r_s = 0.709, n = 28, p < 0.0005). No relationship was found between the SK Scores and the FF Scores at 30 months (Fig. 5.7; r_s = -0.04, n = 31, p = 0.850) or 12 months (Fig. 5.8; n = 29, r_s = -0.06, p = 0.770). No relationship was found between the SK Scores and the HA Latencies at 12 months (Fig. 5.9; n = 32, r_s = 0.137, p = 0.454).

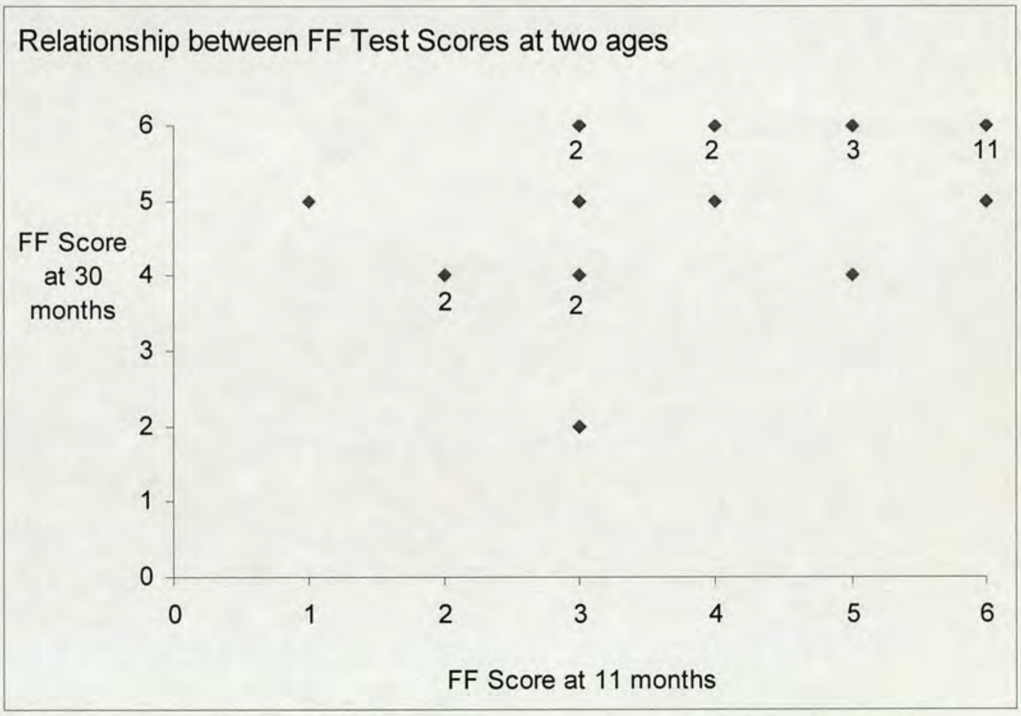


Figure 5.6: Graph showing the relationship between scores from the second Flight-from-Feeder Test carried out on 27 Year 1 heifers at 11 months and 30 months of age. If more than one score lies on the same point on the graph, a number under the point indicates how many scores lie in that position. The score categories range from 1 (most fearful) to 6 (least fearful) and are listed in Table 2.2.

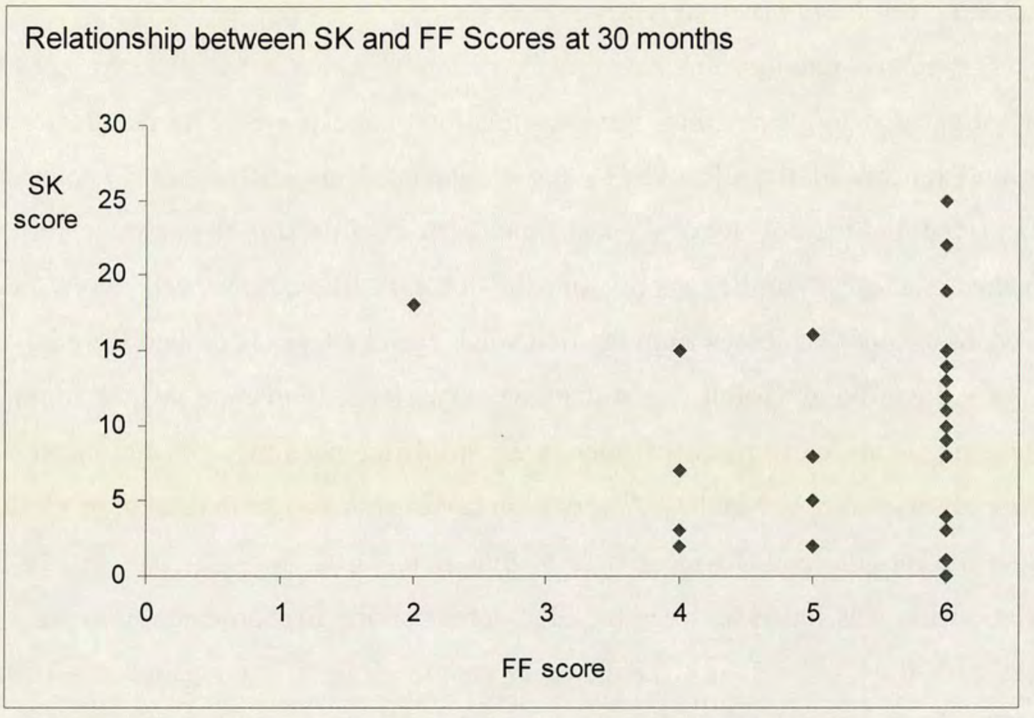


Figure 5.7: Graph showing the relationship between scores from the second Flight-from-Feeder (FF) Test and the total number of steps and kicks (SK) counted in the dairy during the second observations on 31 Year 1 heifers at 30 months. The FF Score categories range from 1 (most fearful) to 6 (least fearful) and are listed in Table 2.2).

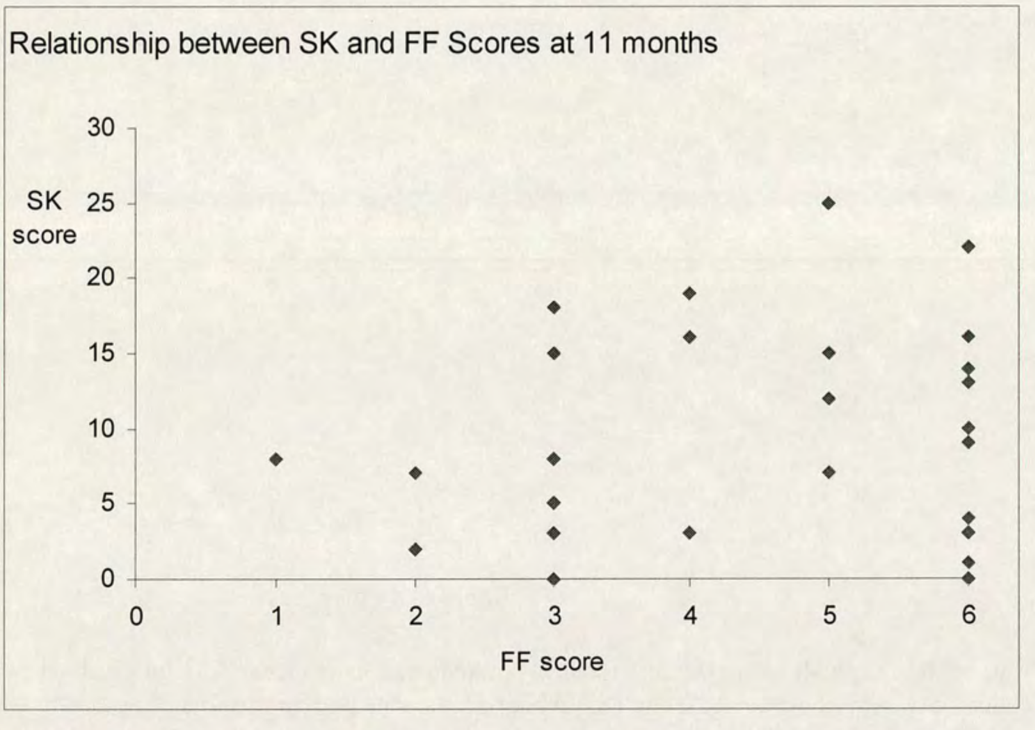


Figure 5.8: Graph showing the relationship between scores from the second Flight-from-Feeder (FF) Test carried out at 11 months, and the total number of steps and kicks (SK) counted in the dairy during observations at 30 months, from 29 Year 1 heifers. The FF Score categories range from 1 (most fearful) to 6 (least fearful) and are listed in Table 2.2.

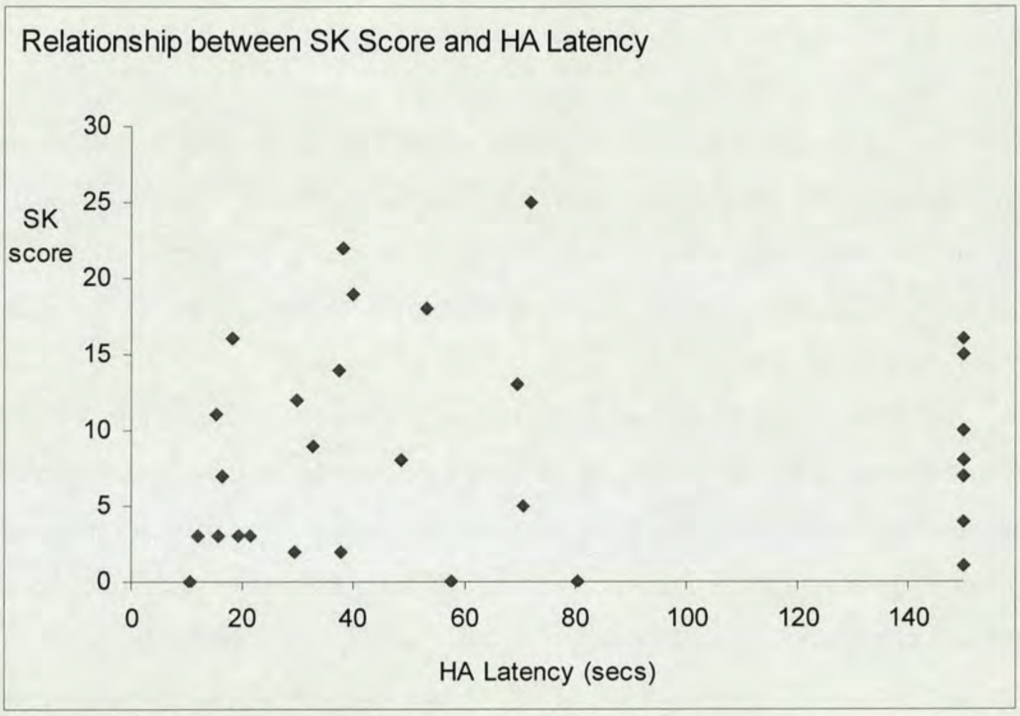


Figure 5.9: Graph showing the relationship between Latencies from the second Handling Test carried out at 11 months, and the total number of steps and kicks (SK) counted during observations in the dairy at 30 months, from 32 Year 1 heifers.

5.3.3. Discussion

This experiment examined whether heifers' behaviours displayed in tests of fearfulness of humans at an early age could be used to predict future behaviour displayed in a dairy parlour. Although measures taken at both ages were highly repeatable, no relationships were found between them.

The FF Tests carried out at 30 months showed a moderate level of repeatability. This result is pleasing, especially considering that the two repeats of the test were carried out from between one and five weeks apart on different animals. However, the high repeatability estimated may be due to the large proportion of very tame animals that were able to be approached closely and touched, meaning that 65 % of the animals scores lay in the top category of the FF Test scale. The heifers' scores were higher in the test than when they were younger, showing that they were less fearful of humans. This may have been due to the increased experience of handling by humans they had encountered by that age. Despite the advantages of this method of carrying out flight tests (discussed in Chapter 3.4), the test is not wholly suitable for distinguishing between heifers that are so tame. Additionally, the method of calculating repeatability may not be as reliable when used on this data with so many values in the top category, and therefore the estimates should be treated with caution. Attention should be paid to the graph showing the mean, minimum and maximum score shown by each heifer (Figure 5.4), to visually assess the extent of the consistency of the score.

A high correlation was seen between these FF Scores and FF Scores previously shown by the heifers at the age of 11 months. Although this result must also be treated cautiously, due to the large number of heifers scoring at the top end of the range at 30 months as discussed above, it implies that the FF Test may provide a trait measure that is stable throughout a long period of a heifers life, between one year and two and a half years of age. This test may therefore be useful to screen young animals to identify those likely to remain very fearful of humans throughout their lifetime. The only other study that has looked at the repeatability of flight tests over long periods of time was carried out by Fisher *et al.* (2000). They found repeatability estimates to be moderate - high when the test was repeated three times at monthly intervals on 20-month old Jersey x Limousin heifers and steers.

The SK Scores showed wide variation between animals and were highly repeatable. This is a very promising result, as the measure used was fairly crude, due to three reasons. Firstly, it summed together the total amount of steps and kicks from three different periods of the milking process. Secondly, the cattle were observed on three days, which differed between heifers in both the time between the heifer calving and the first observation being taken, and also in the number of weeks between the repeat observations. Consequently some of the heifers would have been more used to the process than others at the times at which they were observed. In a previous study, the mean FSK (Flinch, Step, Kick) response of two groups of Friesian or Jersey heifers were monitored over the first 20 weeks after calving. The responses was seen to decline within the first two weeks (Hemsworth *et al.*, 1989), presumably as the heifers habituated to the process. Thirdly, four different stockpersons carried out the milking on different days, and sometimes the handling during the three observation periods was carried out by different people with the same week. The identity of the stockperson handling the animals during milking has previously been found to have an impact on cows' behaviour when being milked (Willis, 1983; Breuer *et al.*, 2000; Hemsworth *et al.*, 2000; Waiblinger *et al.*, 2002).

That a high repeatability was achieved despite these factors, shows that the SK count is a very consistent feature of the animals' behaviour. It is also a quick and easy measure to take, with minimal disruption to the milking procedure involved, so it may prove very useful in studying individual differences in dairy behaviour. SK and FSK responses have been used extensively over the last few years (Rushen *et al.*, 1999; 2001; Breuer *et al.*, 2000; Hemsworth *et al.*, 2000; 2002; Munksgaard *et al.*, 2001; Waiblinger *et al.*, 2002; Van Reenen *et al.*, 2002). Despite this, none of the studies looked at repeatability of the measure, and only one looked at consistency of individual responses across age (Van Reenen *et al.*, 2002). These researchers studied SK counts in Holstein Friesians and found high correlations between counts taken on Day 2 and Day 4 after calving, and moderate correlations between Day 4 and Day 130. Similarly to the present study, most studies did not included the 'flinch' response reported by Willis (1983), and restricted their measurements to steps and kicks.

Although the SK and FF measures at 30 months were both highly repeatable, no relationship was seen between the two measures. It had been hypothesised that, as flight tests and SK behaviour are both thought to be indicative of levels of fearfulness of humans, a relationship between the two behaviours would be evident. Such a relationship between FSK responses in a dairy parlour and flight distances were found in a study looking at individual cows in nine Holstein-Friesian herds (Purcell *et al.*, 1988), but no association was found in two recent studies, one looking at farm averages across 31 Holstein-Friesian herds (Breuer *et al.*, 2000) and one looking at 23 individuals from one Holstein-Friesian herd (Van Reenen *et al.*, 2002).

Relationships between stockperson behaviour in the dairy and FSK responses have been found in some studies, supporting the hypothesis that FSK is at least in part a response to the stockperson (Breuer *et al.*, 2000; Waiblinger *et al.*, 2002). Hemsworth and his colleagues (1987; 1989) carried out two experiments demonstrating that human contact during calving led to a decrease in the FSK response shown by the heifers when milked (1987; 1989). However, Purcell *et al.* (1988) and Van Reenen *et al.* (2002) found no relationship between these variables, leading the authors to suggest that factors other than fear of humans, such as aggressiveness, lameness, vacuum problems of the milking machine, and social pressure from adjacent cows, must play a large part in the FSK response. It seems that factors other than fearfulness of humans played a large part in the responses of the heifers in the present study also. In Section 5.2, the possibility of animals interpreting situations differently depending on their experience and on the particular factors present at the time was discussed. The same test carried out at different ages is likely to measure a different balance of motivations each time. It is possible that the same sort of phenomenon arises between these different studies. For example, where a stockperson displays a high number of negative behaviours towards the cattle in his care, his presence in the dairy is likely to have a large effect on the cows' behaviour, but on a farm where a stockperson handles the animals gently, other factors become more important in explaining the animals' behaviours and motivations.

As no relationship was seen between the dairy SK Scores and FF Scores observed at the same age in the present study, it is not surprising that no relationships were seen between the SK Scores and the FF and HA Test results from the animals at a younger age of 11 months. Again it would appear that different temperament traits underlie the behavioural traits scored using the FF Test and the SK counts. The HA Test measure was shown to be inconsistent when measured across two different ages in the previous section, and it seems it is not a reliable enough measure to be very useful in comparisons with other measures.

To summarise, this experiment showed that FF Scores are highly correlated between the ages of 11 and 30 months, and that SK responses recorded weekly in the dairy are highly repeatable. No relationships were seen between the SK responses and FF Scores measured at 11 or 30 months of age, or between the SSK responses and Latencies measured in the HA Test at 11 months of age.

5.4. Conclusions

The two experiments in this chapter addressed the issue of consistency in temperament traits when measured over long periods of time. The results of both experiments confirmed the existence of behavioural traits in cattle that are consistent over long periods of time, and which are likely to be reflective of underlying temperament traits.

The first experiment described in this chapter examined levels of behaviours assumed to reflect the underlying temperament traits of sociality shown by heifers at four months and then again at 12 months of age. A substantial relationship was seen between levels shown at the two ages. The SS Test could therefore potentially be used to screen animals at an early age to identify those that are likely to show extreme levels of sociality throughout their lifetime. The HA Test, which measured ease of handling, did not provide such a consistent measure.

The second experiment demonstrated that the FF Score is also a behavioural trait of heifers that shows consistency across long periods of time. It is the first experiment that has examined the consistency of response of the FSK response measured in the dairy parlour, and the SK Scores were found to be highly repeatable. However no relationship was seen between the two measures. This, along with findings from other recent studies, suggests that FSK behaviour is not a simple measure of fearfulness towards humans. It is likely that different combinations of factors play a part in the responses seen on different farms. Given the lack of consistency in the results from a number of studies looking at relationships between FSK Scores and either stockperson behaviour in the dairy or flight scores, it seems in future it will be necessary to look to other behavioural measures which may be correlated with the FSK response and which can be easily measured at a young age. It will also be important to study the impact of various factors other than human behaviour on the behaviour of cattle while being milked.

Chapter Six: Genetic and Environmental Effects on Temperament Measures

6.1. Introduction

The aim of the work reported in this chapter was to investigate the genetic basis of temperament traits in cattle. Recent advances in molecular genetics mean that a large number of individual animals can now be genotyped at a large number of DNA markers from throughout the genome. This allows associations between loci present at markers, and quantitative phenotypic trait data measured from the animals, to be found. The locations of markers in the genome are known, and therefore such an association infers the chromosomal region harbouring the quantitative trait loci (QTLs) affecting the trait. The principles of such an analysis are outlined in Chapter 1.7.

Evidence that it should be possible to identify QTL for temperament traits in cattle comes from the large number of studies that have detected QTLs affecting similar traits in mice. As described in Chapter 1.7.3, a series of experiments have independently confirmed a number of QTLs for fear-related traits measured in a range of behavioural tests in mice. The genetic basis of fearfulness traits in cattle may be similar, due to the conservation of homologous genes between species (Flint *et al.*, 1995).

Presently, most of our evidence for the large effects of genetics on temperament traits in cattle comes from differences seen between breeds. Much of this evidence is reviewed in Chapter 1.6. Only two recent studies mapping QTLs for such traits in this species have been carried out to date. Six QTLs were mapped for behaviour measures in offspring from the Canadian Beef Cattle Reference Herd, which was bred from Angus, Belgian Blue, Charolais, Hereford, Limousin and Simmental parent animals (Schmutz *et al.*, 2000). Embryo transfer was used to create this resource herd by producing a population of several full-sib families of purebred and crossbred calves. Levels of agitation and movement of individuals while being held in a weigh scale were measured as an indicator of 'temperament'. Two months later the test was repeated, and the difference between the two scores taken as a measure of 'habituation'. Seven QTLs were found on six chromosomes, five of which affected both measures, and two of which affected 'temperament' only (Schmutz *et al.*, 2001). Work is presently continuing to confirm putative QTLs for behavioural and stress responses found in another resource herd, a Limousin x Jersey herd created in New Zealand and Australia. Preliminary results showed that 'flight distance from an unfamiliar human', measured in the second-generation animals, appeared to be linked to five microsatellite markers (Fisher *et al.*, 2001).

It is therefore an exciting time to look for QTLs for behaviour in a third resource herd. The only way of confirming the existence of QTLs affecting a trait is by independent replication (Lander & Kruglyak, 1995). The RoBoGen herd offered the opportunity to measure similar traits in a different breed cross (Charolais x Holstein). The objective of the experiment on the RoBoGen animals was to carry out a whole-genome scan for QTLs that affect some of the temperament measures taken from the animals that were described in Chapters Three and Four.

Data from two behavioural tests, the Flight-from-Feeder Test (FF Test) and the Social Separation Test (SS Test), were used for the QTL analysis. The FF Test is thought to measure fearfulness of humans, and the SS Test is thought to measure sociality. A large number of animals was required for the QTL analysis, so animals from Years 1, 2 and 3 were tested and the results from all the animals combined for analysis. Genotyping data for the animals was collected separately as part of the parent project.

The experiment is described in this chapter in three parts. The collection of the behavioural data and a summary of the data are described in Section 6.2. Before the analysis of genetic influences was carried out, other factors that may have affected the test results were examined. These were mainly environmental factors associated with the test procedures, and also 'year' and 'sire'. The analyses of these are discussed in Section 6.3. The QTL analysis carried out and the interpretation of the results are described in Section 6.4.

6.2. Behavioural Data Collection

6.2.1. Methods
6.2.2. Summary of data

6.2.1. Methods

The two behaviour tests were carried out on the F_2-generation animals born in Years 1 - 3. FF Tests were carried out on the 181 bulls and 197 heifers, and SS Tests were carried out on the heifers only. The test procedures used are described in Chapter 2.3.2 and 2.3.3. Briefly, the measure taken in the FF Test was a Flight Score, which was thought to represent the level of fearfulness of the animal towards humans. The FF Score scale ranged from 0 - 6, 0 signifying high fearfulness and 6 indicating low fearfulness (see Table 2.2). The measure taken in the SS Test is the duration of time spent Walking, Running or showing Escape behaviours (WER). This measure is thought to be a measure of sociality, a high WER Score indicating high sociality.

The test schedules followed each year are described in Chapter 2.4. Each schedule included additional tests that were carried out as part of the experiments described in Chapters Three to Five. In Year 1, FF, SS, Novel Object (NO) and Handling (HA) Tests were carried out on the bulls and heifers, and the data obtained were examined for inter-animal variation and intra-animal repeatability (as described in Chapter Three) and for inter-test validity (as described in Chapter Four). These analyses showed the FF and SS Tests to give the most useful results in terms of measuring repeatable traits that were stable over long periods of time. However, the two tests remain to be validated with other tests, and so the interpretations of the behaviours as indicating underlying temperament traits of fear of humans and sociality should be treated with caution.

The collection of data from the FF and SS Tests was therefore continued on further animals in Year 2 and Year 3 for QTL analysis. The FF Tests were carried out two or three times on each animal, on the same day or up to eight days later. The SS Tests were carried out two or three times on consecutive days. The bulls were approximately eight months of age when the FF Tests were carried out, and the heifers were approximately 10 – 12 months of age. The number of animals tested each year is detailed in Table 6.1.

Before the data from the three years was combined for the QTL analysis, summary statistics for each score were calculated from the data collected in each year, in order to check that similar values and patterns were seen between years. The analyses carried out were similar to those carried out on the data from Year 1 in Chapter Three. The median and range of each measure was calculated for the first and second repeats of each test. Principal Components Analysis (PCA) was used to interpret behaviours seen in the SS Test. Further details of these statistical methods are given in Chapter 2.4. Paired t-tests were used to test for differences in the mean response across all of the animals between the first two test repeats. Finally, repeatability of each measure was estimated using a ratio of the variance components calculated using REML.

In Chapter Three it was concluded that the data from the second repeat of each test was likely to be the most useful measure of behaviour resulting from one main motivation, as the novelty of the test situation appeared to have a large effect on the responses to the first test of each type. Consequently the data from the second test repeats was used for the genetic analysis. Much of the description in this chapter therefore focuses on the data from the second repeats of the tests.

6.2.2. Summary of Data

6.2.2.1. Summary of bull data - FF Test

The statistics described above were calculated for the FF Tests carried out on the bulls and are summarised in Table 6.1. A higher FF Score indicated a less fearful animal. The median Test 1 scores get lower across the three year groups, implying that the bulls in the later two years were generally more fearful than those from Year 1. The median and range of the bulls' FF Scores in Test 2 is the same for Years 1 and 2, with a median of 3 being the central score in the scale, and a range covering 6 of the 7 categories of the scale. In the third year the median increases by 1, implying that the Year 3 bulls may be less fearful, and the range covers all seven possible categories. Comparison of the overall mean response of the animals between the first and second tests showed no change across the two test repeats in Year 1 or 2, but an increase in Year 3. This implies that the bulls in Year 3 became less fearful after the first test.

The repeatability of the FF Scores was high in Year 1 (see Chapter 3.3.1), and it was on the basis of this result that the test was continued in future years. However, in Year 2 the test results were not found to be repeatable, and in Year 3 the repeatability estimate was moderate at 0.31 ± 0.07 (see Table 6.1). Why there should be such big differences in repeatability of the trait between years is not immediately apparent. Figure 6.1 shows the mean FF Score from the two tests for each animal, along with the minimum and maximum scores achieved. The distribution of scores looks very similar between each of the three years. Figure 6.1b shows that there was a large number of bulls in Year 2 that differ substantially in their mean score, and whose minimum and maximum score bars do not overlap, giving the impression of a high repeatability.

It was suspected that the repeatability estimate may be affected by the number of bulls that show a difference of more than 1 between their two test scores. To investigate whether this was the case, the twelve bulls that showed a mean score of 3 and a minimum and maximum score of 2 and 4 respectively, were changed to having scored 3 on both tests, i.e., now having a minimum and maximum score of 3, and the repeatability analysis was repeated. The value obtained was moderate at 0.19 ± 0.12, showing that this small change to the data had quite an effect on the repeatability value obtained. This implies that repeatability calculations do not reflect accurately the pattern of the FF Test Score data, probably because the scale is so limited and the data is categorical. The graphs in Figure 6.1 show repeatable differences between individuals in all years, suggesting that the data is suitable as a trait measure.

6.2.2.2. Summary of heifer data - FF Test

The median FF Scores shown by the heifers were higher than those shown by the bulls each year, reflecting their lower fearfulness of humans (Table 6.1). This difference was probably due to the differences in the rearing environment experienced by the two sexes. The median scores did not decrease across successive years, as seen in the bulls' scores, apart from a lower score seen in the second test carried out on the Year 3 animals. In Years 1 and 3, the heifers showed a decrease in the score between Tests 1 and 2, implying that they became more wary of the observer after the FF Test had been carried out once. Over all years, six of the seven categories are scored in the tests. The repeatability estimate was high in Year 1 ($r = 0.58 \pm 0.09$), moderate in Year 2 ($r = 0.40 \pm 0.10$), but not repeatable in Year 3. Similarly to the bulls, these large differences between years were unexpected. The median scores obtained by the heifers are plotted in Figure 6.2, along with minimum and maximum bars. The scores appear to be repeatable in all three years. The only obvious difference between the data in Year 3 seems to be that only one heifer achieved the maximum score once, making the range of data smaller. To explore whether the smaller range was making the difference to the repeatability scores, a small change was made to the data from Year 3 and the repeatability recalculated. When fourteen median scores of 5 (shown in Figure 6.2c) were changed to the maximum value of 6, the repeatability estimate went up to 0.432 ± 0.1. This illustrates that perhaps calculating repeatability does not give a reliable summary of ordinal, limited-scale data such as this.

6.2.2.3. Summary of heifer data - SS Test

In Chapter Three, a PCA was carried out on the behaviours shown in the SS Test. From this, 'WER' was selected as a measure of sociality, as explained in Chapter 3.3.2. PCA was carried out again on the data from the heifers from all years, to check if the same pattern was seen in the data. All heifers that were tested in the second repeat of the test were included ($n = 189$). Figure 6.3 shows the first two components of the analysis plotted as a graph. The graph shows a similar pattern of clustering of behaviours as seen in the Year 1 heifers (see Figure 3.3b). 32.1 % of the variation is represented by Component 1, which contrasts individuals with long durations of W, R, E and a high frequency of vocalise (V), with those with long durations of time spent sniffing, licking or rubbing the wooden boards (B). This graph reaffirmed the decisions drawn from the Year 1 data, that WER is a useful trait measure to use from the heifers.

Table 6.1 shows that the two median WER Scores for the two repeats of the tests are higher in Year 2 than in Years 1 and 3, implying that the Year 2 heifers showed higher levels of sociality. Comparisons of the mean difference in responses between Tests 1 and 2 using Paired t-tests showed there was no consistent change in response after the first test. The ranges are very wide and are fairly similar between years. A high degree of repeatability was found in the tests responses in all three years ($r = 0.59 - 0.77$), and when the data from all three is combined ($r = 0.75 \pm 0.03$). The median WER Score for each animal and the minimum and maximum scores shown by each are shown in Fig. 6.4.

Table 6.1: Summary statistics of scores from the FF Tests carried out on the bulls, and FF and SS Tests carried out on the heifers. FF Score = score from the Flight-from-Feeder Test (see Table 2.2). WER Score = duration of test time spent Running, showing Escape behaviour or Walking in the Social Separation Test. The median score and range of scores is presented for the first and second repeats of each test, Tests 1 and 2. Changes in mean responses between Tests 1 and 2 were detected using Paired t-tests. Repeatability values between Tests 1 and 2 were calculated using ratios of variance components calculated using REML. T = t statistic, NS = $p > 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Test Measure	Animals	Year Group	Median and Range of Test 1 data	Median and Range of Test 2 data	Change in mean response between Tests 1 and 2	Repeatability ± standard error
FF Score	Bulls	Year 1 (n = 48)	4 (1 - 6)	3 (1 - 6)	No change, T = 1.11 NS	0.47 ± 0.11 ***
		Year 2 (n = 64)	3 (1 - 5)	3 (1 - 6)	No change, T = -1.54 NS	0.06 ± 0.12 NS
		Year 3 (n = 64)	2 (0 - 6)	4 (0 - 6)	Increase 0.823 T = -4.94 ***	0.38 ± 0.11 **
		All years (n = 176)	3 (0 - 6)	4 (0 - 6)	Increase 0.353, T = -2.86 **	0.31 ± 0.07 ***
FF Score	Heifers	Year 1 (n = 54)	5 (2 - 6)	5 (1 - 6)	Decrease 0.491, T = 3.40 **	0.58 ± 0.09 ***
		Year 2 (n = 72)	5 (2 - 6)	5 (2 - 6)	No change, T = 1.63 NS	0.40 ± 0.10 ***
		Year 3 (n = 69)	5 (2 - 5)	4 (1 - 6)	Decrease 0.515, T = 3.81 ***	0.12 ± 0.12 NS
		All years (n = 195)	5 (2 - 6)	5 (1 - 6)	Decrease 0.393, T = 5.00 ***	0.43 ± 0.06 ***
WER Score	Heifers	Year 1 (n = 54)	54.8 (0.0 - 249.8)	38.8 (0.0 - 250.2)	No change, T = 0.44 NS	0.59 ± 0.09 ***
		Year 2 (n = 72)	173.8 (8.3 - 262.6)	174.1 (16.8 - 262.8)	No change, T = 0.78 NS	0.59 ± 0.08 ***
		Year 3 (n = 67)	83.3 (9.1 - 239.6)	79.8 (10.2 - 238.3)	No change, T = 0.94 NS	0.77 ± 0.05 ***
		All years (n = 193)	101.0 (0.0 - 262.6)	102.7 (0.0 - 262.8)	No change, T = 0.20 NS	0.75 ± 0.03 ***

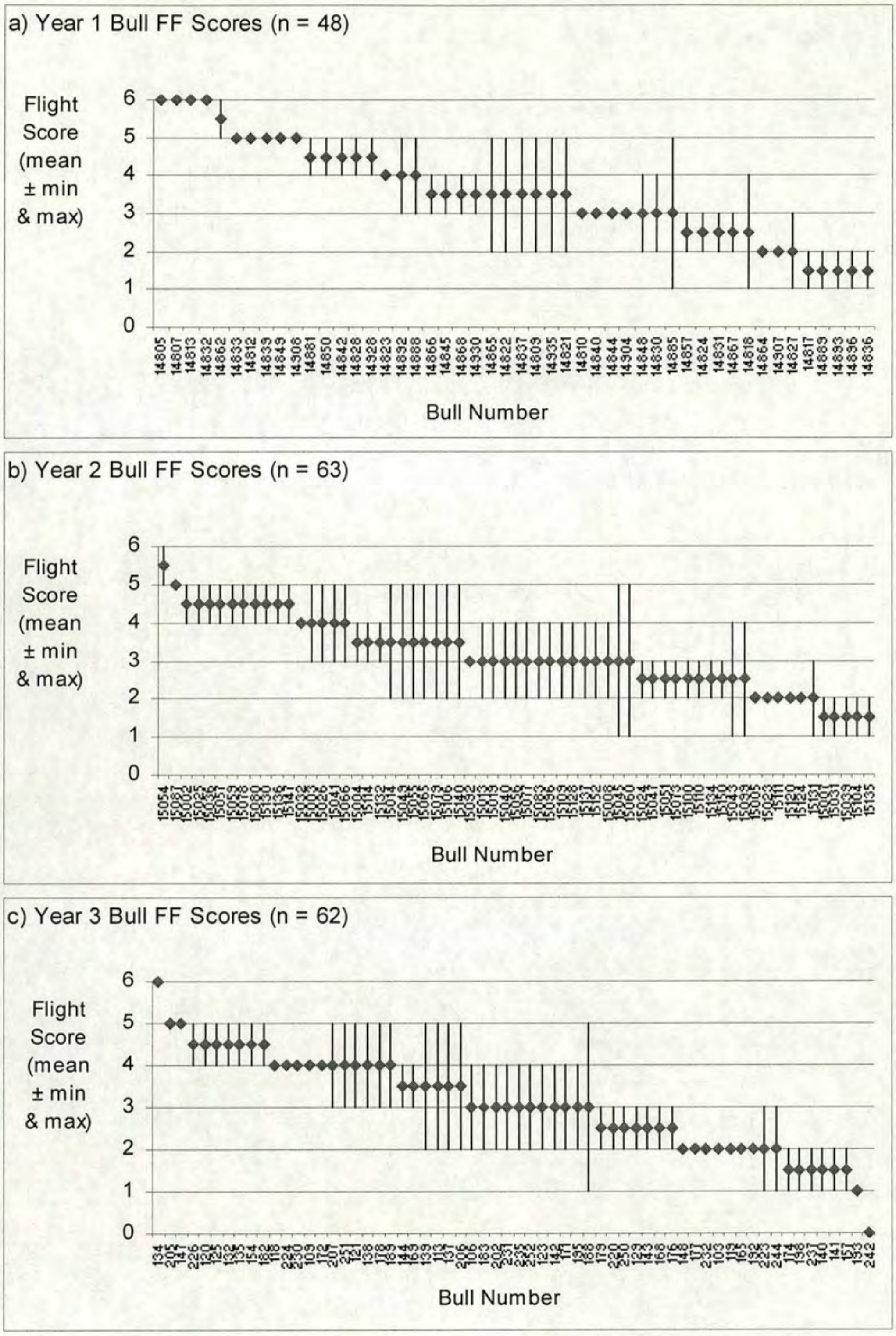


Figure 6.1: Flight-from-Feeder Scores for the bull calves, in order of mean score. The scores are means from two test repeats. Bars show the minimum and maximum scores gained by each animal. Key to scores: 0 'would not approach feeders while observer present', 1 'move away within 2.00 - 1.25m', 2 'move away within 1.00 - 0.25m', 3 'move away at 0m', 4 'move back as touch is attempted', 5 'move back when touched', 6 'didn't move back when touched'.

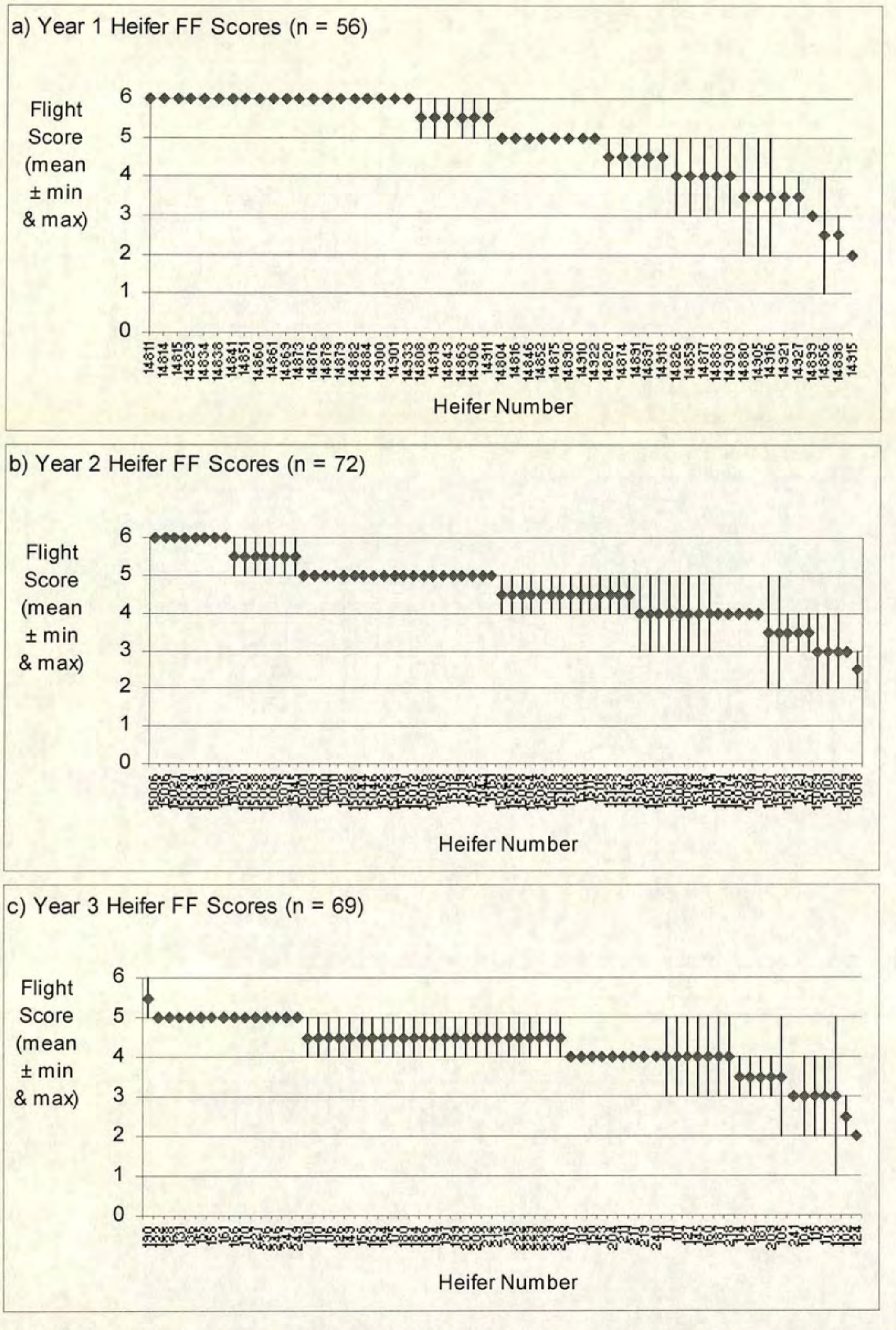


Figure 6.2: Flight-from-Feeder Scores for the heifer calves, in order of mean score. The scores are means from two test repeats. Bars show the minimum and maximum scores gained by each animal. Key to scores: 0 'would not approach feeders while observer present', 1 'move away within 2.00 - 1.25m', 2 'move away within 1.00 - 0.25m', 3 'move away at 0m', 4 'move back as touch is attempted', 5 'move back when touched', 6 'didn't move back when touched'.

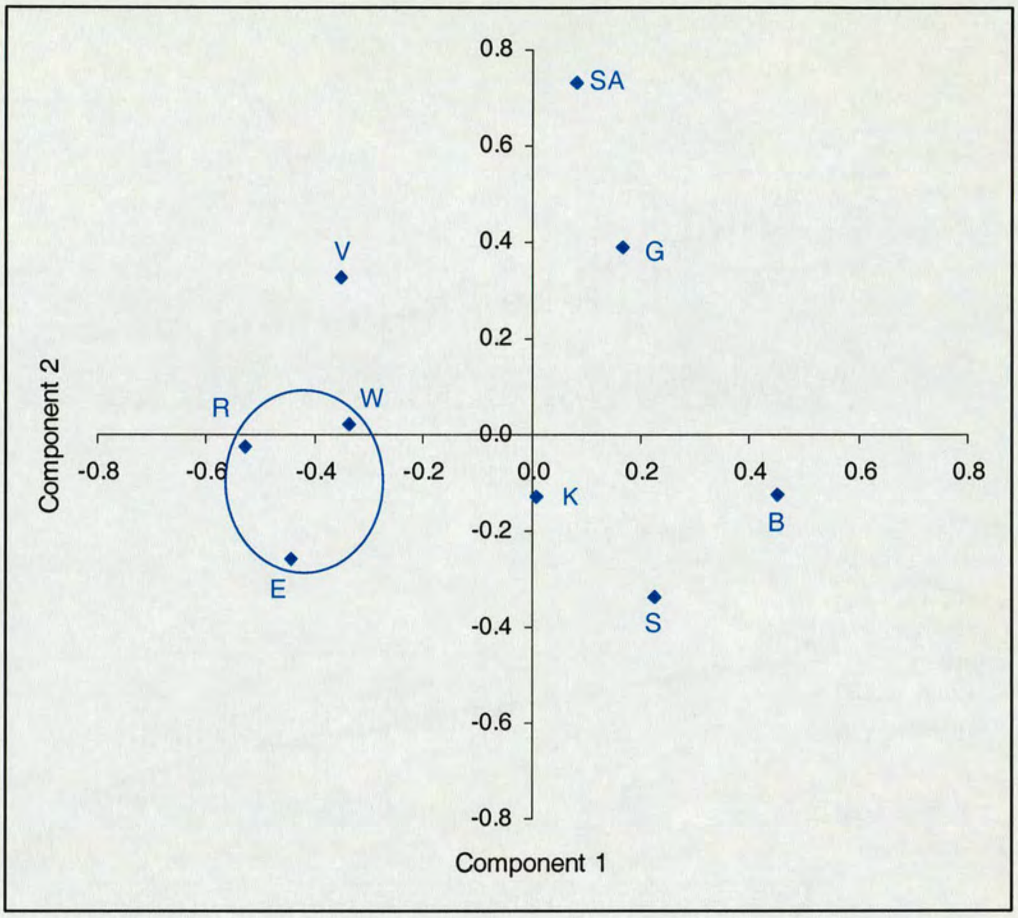


Figure 6.3: A Principal Components Analysis of data from the second SS Test carried out on 189 heifers from all three years. The behaviours measured are durations of B = stand and sniff/lick/rub boards, E = escape, G = gambol, K = kneel, R = run, S = stand occupied, SA = stand alert, W = walk, and a frequency V = vocalise.

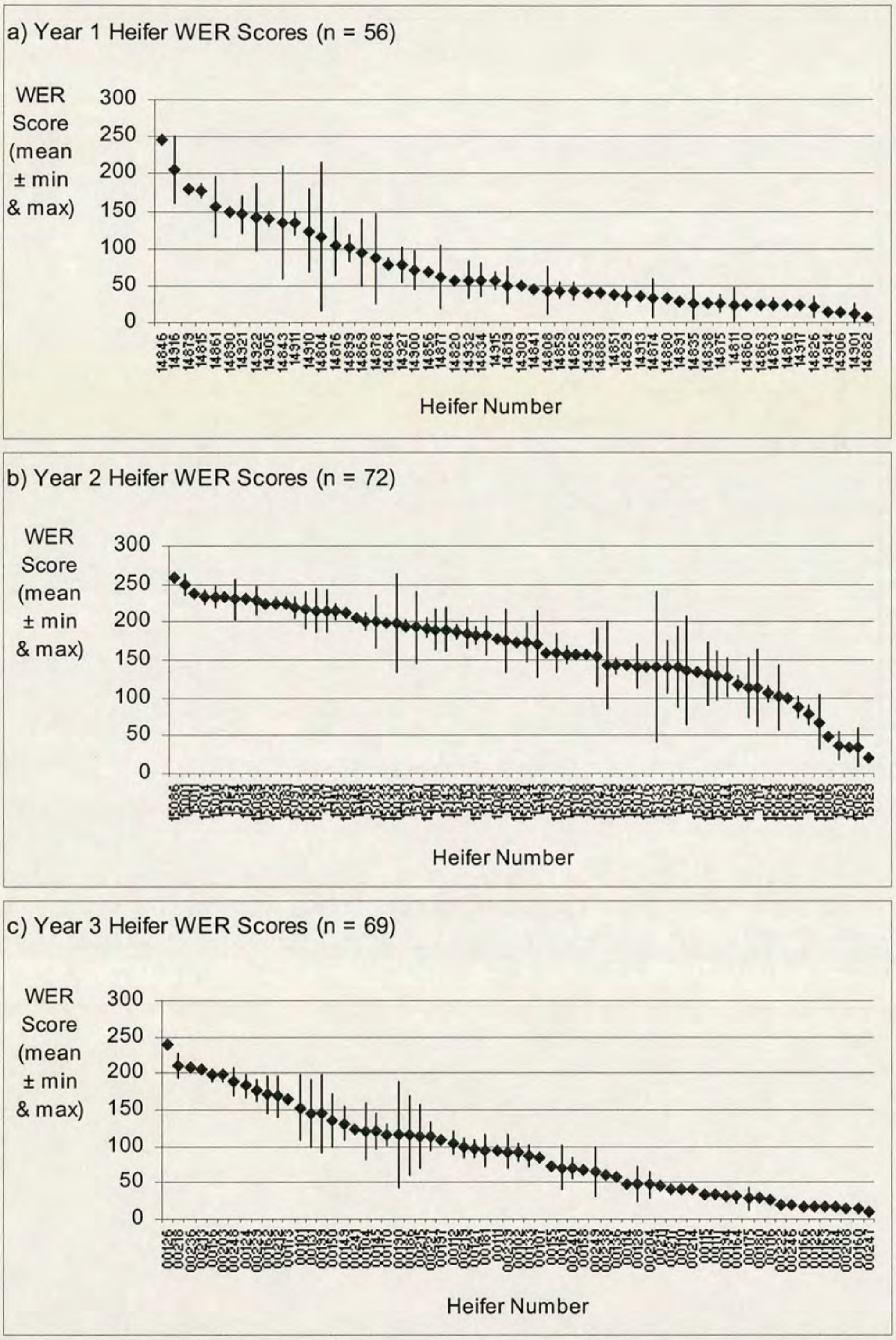


Figure 6.4: WER Scores for the heifer calves, in order of mean score. WER = the duration of test time spent Walking, showing Escape behaviour, or Running in the Social Separation Test. The mean scores are from two test repeats.

6.2.2.4. Discussion

These analyses examined the suitability of the behavioural data from the animals from the three years for use as phenotypic trait measures for QTL analysis. The FF Scores showed wide variation between the animals in each of the three years, and when examined graphically, appear to be repeatable in each year. Despite this, the estimates of repeatability indicated that the scores from the bulls in Year 2 and the heifers in Year 3 were not repeatable. When investigated it appeared that the repeatability estimates of the FF Score were very sensitive to a small change in the range of scores obtained, probably because the scale is so limited. When the data from all the years was combined, the repeatability estimate was moderate. These results indicate that the method or scale of scoring should be altered in future tests, so that the data produced is more amenable to such statistical analysis.

The mean WER Scores from the heifers also showed wide variation between animals, and appeared to be highly repeatable from the graphs. This was backed up by high repeatability estimates for each year, and when the data was combined.

The Year 3 bulls showed lower median FF Scores than the bulls in the other two years, and the Year 2 heifers appeared to have higher median levels of sociality than the other heifers. The possible effect of Year on the scores is investigated in Section 6.3.

In conclusion, both the FF and SS Test scores from all three years are suitable for using as trait measures in the QTL analysis.

6.3. Analyses of Fixed Effects

6.3.1. Methods of REML Analyses
6.3.2. Results and Discussion

6.3.1. Methods of REML Analyses

REML calculations were carried out to investigate whether factors other than differences between the individual animals caused variation in the FF and SS Test Scores, and should therefore be included in the QTL analysis as fixed effects. REML was used to attribute the total variation seen in the behavioural data into different sources (or components) of relative magnitude (see Chapter 2.5.2). Environmental factors associated with the test procedures were investigated as sources of variation; 'year', 'housing group', 'test order' and 'age'. Genetic influences were also investigated by looking at the effects of 'sire' and 'breed group'. These factors are explained further below. FF Test Scores from the second repeat of the FF Test were examined in the bulls and the heifers, and WER from the second repeat of the SS Test in the heifers. As the FF data used an ordinal scale and was therefore only approximately normally distributed, the results were interpreted with caution.

As there were a large number of factors, they were split into two separate REML analyses: one with effects due to the testing procedure and housing arrangements, and one with effects due to differences between the animals. In each model, 'Animal ID' was used as the residual factor. 'Year' defined whether the animal was born in Year 1, 2 or 3. 'Housing Group' defined which group the animal was housed in. There were 13 groups of bulls; A - D in Year 1, E - H in Year 2 and I - M in Year 3. There were 16 groups of heifers; A - D in Year 1, E - J in Year 2 and K - P in Year 3. 'Test Order' described the order in which the animals were tested on the afternoon of the test in question, and applied to the SS Tests but not the FF Tests. The order was classed into three categories; 1 = tested 1st - 5th, 2 = tested 6th - 10th, 3 = tested 11th - 15th). 'Breed Group' indicated whether the animal was an F_2, a Charolais backcross, or a Holstein backcross. 'Sire' gave the identity of the animal's sire, of which there were eight F_1 and four Charolais sires. 'Age' gave the animal's age, and this measure was also divided into three or four classes. The age classes differed between the FF and SS Tests, as they were carried out at different times (as described in Chapter 2.4).

The analyses were carried out using REML procedures in Genstat for Windows (Release 4.21, Rothamsted Experimental Station, 2001). Variance components for each factor were obtained and examined. Significance levels were investigated initially by sequentially omitting factors with small variance components before reinstating them in the models (Genstat 5 Release 3 Reference Manual, 1993, p. 571). Any component that was twice as large as its standard error was investigated further for use as a potential fixed effect in the QTL analysis.

6.3.2. Results and Discussion

6.3.2.1. Fixed Effects on Bull FF Scores

Environmental factors associated with the test procedures were investigated in the first analysis on the scores from the second FF Test carried out on the bulls. The variance components of Year and Housing Group were calculated, and these are shown in Table 6.2. Year and Housing Group explain none of the variance in the data, but a large variance is seen between the bulls' scores. Table 6.2 also shows the variance components calculated for Breed Group, Sire and Age. Age was classed into four categories; 1 = 211 - 250 days, 2 = 251 - 290 days, 3 = 291 - 330 days and 4 = 331 - 370 days. Breeding Group and Age account for none of the variance seen in the FF Scores, and Sire accounts for a negligible amount. Again, a large amount of variance is seen between the individual bulls, which is not explained by the factors investigated.

Table 6.2: Summary of REML analyses of factors which may affect FF Test 2 Scores in the bulls. Estimated variance components and standard errors were calculated for each factor. The factors are listed in the order they were entered into the model.

Scores	REML analysis	Factor	Variance Component ± s.e.
FF	Environmental Factors	Year	0.000 ± 0.000
		Housing Group	0.000 ± 0.000
		Bull ID (residual term)	1.851 ± 0.200
	Animal Factors	Breed Group	0.000 ± 0.000
		Sire ID	0.040 ± 0.076
		Age	0.000 ± 0.000
		Bull ID (residual term)	1.815 ± 0.203

6.3.2.2. Fixed Effects on Heifer FF Scores

The same factors were investigated for the FF Scores from the second FF Test in the heifers. In the first analysis Year and Housing Group were used as factors. The REML output is shown in Table 6.3. The analysis shows that Year and Housing group explain a negligible amount of variance in FF Score. A substantial amount of variance is seen between the heifers. A further analysis detailed in Table 6.3 examines the effects of Breed Group, Sire and Age. Age was divided into four classes; 1 = 251 - 290 days, 2 = 291 - 330 days, 3 = 331 - 370 days and 4 = 371 - 385 days. Breed Group and Sire ID explain none of the variation in the test scores, and Age a negligible amount. Again a large amount of variance is seen between the individuals' FF Scores.

Table 6.3: Summary of REML analyses of factors which may affect FF Test 2 and SS Test 2 Scores in the heifers. Estimated variance components and standard errors were calculated for each factor.

Scores	REML analysis	Factor	Variance Component ± s.e.
FF	Environmental Factors	Year	0.116 ± 0.143
		Housing Group	0.014 ± 0.051
		Heifer ID (residual term)	1.487 ± 0.158
	Animal Factors	Breed Group	0.000 ± 0.000
		Sire ID	0.000 ± 0.000
		Age	0.089 ± 0.109
		Heifer ID (residual term)	1.492 ± 0.176
WER	Environmental Factors	Year	2216 ± 2377
		Housing Group	277 ± 364
		Testing Order	642 ± 430
		Heifer ID (residual term)	3868 ± 455
	Animal Factors	Breed Group	30 ± 173
		Sire ID	0 ± 0
		Age	404 ± 423
		Heifer ID (residual term)	5776 ± 668

6.3.2.3. Fixed effects on Heifer WER Scores

The factors investigated in the WER Scores from the second SS Test on the heifers were Year, Housing Group and additionally Test Order. The variance components listed in Table 6.3 show that Year, Housing Group and Test Order explain a negligible amount of the variance seen in the data. Animal ID again has substantial variance. A second analysis carried out investigated the effects of Breed Group, Sire and Age. Age was divided into classes: 1 = 301-340 days, 2 = 341-380 days, 3 = 381-420 days. As shown in Table 6.3, Sire had no effect, Breed Group and Age had negligible effects, and again a large amount of variation is seen between the heifers.

6.3.2.3. Discussion

None of the environmental factors investigated had an impact on the FF Scores from the bulls and heifers, or the WER Scores from the heifers. Substantial variation in test score was seen between the animals in each case, none of which was explained by any of the factors in the models. This implies that both tests are robust against the effects of the year the animals were born, the group the animals were housed in, the age of the animal at testing and the day on which the test was carried out. This finding that age has no effect is of particular interest, as there were differences of up to four months between the bulls when FF Testing took place, and differences up to five months between the heifers when their first tests were carried out (see Chapter 2.4).

It is surprising that no effects of breed groups were seen, as the two founder breeds are expected to differ in these traits. Previous differences in flight distance have been reported between Charolais and Friesian cows (Murphey *et al.*, 1980). This may imply that a QTL analysis on this population would have a low power (probability of detection) for QTLs for these traits. Neither were Sire effects found. The segregation of genes does not occur until the F_2 generation, so differences in the traits are not expected between F_1 sires. Differences would be expected between the F_1 sires and the Charolais sires, however most of the sires were F_1s.

In conclusion, none of the factors examined need to be considered as fixed effects when the test scores are analysed for QTLs. However, as the FF Scores were from a limited, ordinal scale, care should be taken not to rely too heavily on these results. It was thought prudent to include 'Year' in the QTL analyses as a fixed effect. The FF Scores from the bulls and the heifers were analysed together to scan for QTLs, to obtain a large a sample size as possible, and so sex was also included in the QTL analysis model as a fixed effect.

6.4. QTL Analyses

6.4.1. Methods

6.4.2. Results and Discussion

6.4.1. Methods

6.4.1.1. Phenotypic Data

The data described above in 6.2.2 from the FF Tests in the F_2 and backcross bulls and heifers, and the WER Scores from the SS Tests in the heifers, were collated for QTL analysis. Scores from the second repeat of each test were used. The scores for 23 of the animals phenotyped were taken out, as the genotyping data required for analysis was not yet available for them. This left 343 bulls and heifers with FF Scores and 177 heifers with WER Scores.

6.4.1.2. Genotypic Data

The genotyping procedure carried out as part of the parent project is described in Chapter 2.6.2. Briefly, markers to be used were selected from publicly available genetic linkage maps, and tested for their information content in the RoBoGen population. DNA was then obtained from all the grandparents, parents and offspring animals and sent off to a commercial laboratory in the USA for genotyping at the selected markers. The genotyping data was received electronically.

The genotyping was held up in Year 2 of the project due to import restrictions put in place by the USA due to the Foot-and-Mouth outbreak in the UK. This led to a long delay in obtaining the genotyping results. At the start of this analysis, the genotyping data had only just been received from the laboratory. It contained many errors which had to be corrected before running the QTL analysis. The errors came from two sources: incorrect reading of allele sizes from the gels, and incorrect recording of the pedigree information. Additionally, data were missing for some animals whose DNA was missed when each batch was sent off for analysis, or where technical problems had been encountered in the laboratory. As it became clear that a complete genotyping data set would not be available for analysis for several months, initial plans to carry out a genome scan on all 29 pairs of autosomes and two sex chromosomes were scaled down. It was decided to concentrate the time available on correcting the errors for Chromosome 1 and then running the QTL analysis on that chromosome only.

F_2 or backcross offspring that had not been genotyped were excluded, leaving the numbers mentioned above in 6.4.1.1. The two allele sizes for each animal at each of the chromosome 1 markers were checked against the pedigree for errors in Mendelian inheritance, using the 'gtcheck' option in the CRI-MAP Unix program (Green *et al.*, 1990). A list of genotyping inconsistencies was produced. These were examined, and where the correct alleles sizes could not be determined through inheritance checking, the genotypes were temporarily set to unknown. Pedigree inconsistencies were also examined. Where sire or dam errors could not be resolved, temporary made-up parent names were entered to allow the offspring involved to be accepted by the program. This checking and correcting procedure was repeated until all errors were resolved or set to unknown. The data was then recoded and the 'build' option in CRI-MAP was used to check approximate map distances between and order of markers. These were similar to those obtained from the IBRP and MARC consensus maps (see Chapter 2.6.2). As more data had been used to construct the published maps from which the consensus maps had been assembled, the consensus map distances were used for the analysis. Finally, the data was then ready to run in QTL Express.

It is anticipated that almost all of the errors involving animals with missing or incorrect genotypes or unknown parents will eventually be resolved with further checking of allele readings and genotyping of individuals. However, at the time of doing the analysis reported here, the data set was incomplete.

Additional help was provided which enabled corrected data from chromosomes 9, 14 and 15 also to be made available for analysis. These chromosomes were selected as they were among those identified as harbouring QTLs for temperament traits in a previous study (Schmutz *et al.*, 2001). The markers on each of the four chromosomes and the map distances used are shown in Table 6.4. The percentage of genotyping data available from the population for each of the chromosomes is listed in Table 6.5, and was at least 90 % for each.

Table 6.4: A list of the markers and marker distances used for the QTL analysis. The marker distances were obtained from the USDA and MARC consensus maps (see Chapter 2.6.2).

Chromosome Number	Marker Number	Marker Name	Distance between consecutive markers (cM)
1	1	TGLA49	0
	2	BMS4017	32.9
	3	TGLA57	11.4
	4	INRA128	14.8
	5	BM864	27.2
	6	CSSM19	20.1
	7	BMS4044	20.4
9	1	ETH225	0
	2	BM2504	17.1
	3	UWCA9	19.7
	4	MM12E6	32.5
	5	INRA84	6.9
14	1	CSSM66	0
	2	RM11	24.7
	3	PZ271	33.1
	4	BM4513	1.7
	5	BM2934	3.9
15	1	BR3510	0
	2	JAB1	19.8
	3	BMS2684	14.1
	4	IDVGA10	16.7
	5	BMS429	41.8

6.4.1.3. QTL Analysis

The 'QTL Express' analysis program (Seaton *et al.*, 2002) was used to look for linkage between each trait and each chromosome in turn, and is described in Chapter 2.6.3. The program update that allows backcross animals to be included in the analysis was still under development. Bugs found in this version meant that a further two animals had to be taken out of the genotype file of chromosomes 9 and 15 to allow them to run, accounting for the slightly reduced numbers of animals reported for these two chromosomes in Table 6.5.

A one-QTL analysis using a model with additive and dominance effects was carried out with a 1 cM step size. The model used for analyses of FF also included fixed effects of year and sex, and the model for WER included the fixed effect of year. Permutations with 1000 iterations were performed to calculate the single-position and chromosome-wide significance thresholds for the F-value of each analysis (see Chapter 2.6.3).

Chromosomes that showed significant evidence at the single-point level for QTLs affecting either trait were examined further. In each FF analysis, a QTL x sex interaction was tested to investigate whether the effect differs between the two sexes.

6.4.2. Results and Discussion

The scan of the four chromosomes for QTLs affecting FF Score revealed two locations, one on Chromosome 14 and one on Chromosome 15, which were significant at the single-position level. A QTL for WER was found on chromosome 14, which was significant at the single-position level. The F-ratios and locations of these putative QTLs are listed in Table 6.5, along with estimates of additive and dominant effects. Tests were carried out to investigate interactions of sex and year with the QTLs. Neither of the traits showed significant interaction with either factor when models with an interaction were compared with models without interactions (Table 6.6).

Table 6.5: Results of a four-chromosome scan for QTLs for temperament traits, using a model with additive and dominant effects. FF Scores are from the second repeat of the Flight-from-Feeder Test. WER are total duration of test time spent Walking, Running or showing Escape behaviour from the second repeat of the Social Separation Test. Significance levels of the F-ratios: NS = $p > 0.05$ at single-position permutation with 1000 iterations, single position = $p < 0.05$ at single-position permutation with 1000 iterations, chromosome = $p < 0.05$ at chromosome-wide permutation with 1000 iterations.

Chromosome (% genotypes)	Trait	Animals	No. of animals	Fixed effects	F-ratio	Significance level	QTL location (cM)	Additive Effect ± s.e.	Dominant Effect ± s.e.
1	FF	heifers + bulls	343	year + sex	1.66	NS	-	-	-
(92.7)	WER	heifers	177	year	1.17	NS	-	-	-
9	FF	heifers + bulls	341	year + sex	2.18	NS	-	-	-
(91.8)	WER	heifers	175	year	2.61	NS	-	-	-
14	FF	heifers + bulls	343	year + sex	3.20	single position	58	-0.12 ± 0.14	0.51 ± 0.20
(90.1)	WER	heifers	177	year	3.76	single position	12	25.14 ± 11.40	-34.27 ± 20.39
15	FF	heifers + bulls	341	year + sex	3.43	single position	13	0.032 ± 0.15	-0.67 ± 0.26
(93.0)	WER	heifers	175	year	2.41	NS	-	-	-

Table 6.6: Results of fitting interactions with sex and year compared to the model with no interactions with significant results. NS = p > 0.05.

Chromosome	Trait	Interaction	F-ratio	Significance level
14	FF	Sex	1.26	NS
		Year	0.69	NS
	WER	Year	2.58	NS
15	FF	Sex	1.06	NS
		Year	0.98	NS

Although these three possible QTL locations are of interest, their significance should be treated with caution. At the single-position level, many significant QTLs are likely to be identified by chance due to the number of tests being carried out (type 1 error). Hence, firm conclusions about the locations and effects of these QTLs must be delayed until a genome-wide scan can be carried out on the data when the genotyping data set is complete. It is likely that some, but not all, of the QTLs identified here will exceed the chromosomal or genome-wide significance thresholds used when this analysis is possible.

Nevertheless, the locations identified are interesting, and can be speculatively compared to the locations of QTLs found in a previous genome scan for temperament traits in cattle. Seven QTLs for two behavioural measures were identified in the Canadian Beef Cattle Reference Herd (Schmutz *et al.*, 2001). 'Temperament' was inferred by agitation and movement of each animal whilst held in a weigh scale, and the difference between two temperament scores taken two months apart was called a 'habituation' measure. Five QTLs each affecting temperament and habituation were found on chromosomes 1, 5, 9, 11 and 15, and two additional QTLs affecting temperament alone were found on chromosome 14 (Table 6.7). The locations of the QTLs for temperament identified on chromosome 14 in the Canadian herd, 19 and 35 cM, are similar to the locations where QTLs have been speculatively identified in this study, for WER at 12 cM, and FF at 58 cM. The putative QTL found here for FF on chromosome 15 at 13 cM lies very close to the same position as the QTL found in the Canadian herd at 12 cM on chromosome 15. This gives very strong support for the QTL identified here on chromosome 15, which is therefore worthy of further investigation. The QTLs identified on chromosome 14 are also likely to be of great interest.

Table 6.7: QTLs for temperament traits in cattle identified in the study carried out by Schmutz and her colleagues (2001) and in the present study, listed by chromosome. 'Temperament' was inferred by agitation and movement of each animal whilst held in a weigh scale, and the difference between two temperament scores taken two months apart was called a 'habituation' measure. In the present study, FF was the trait measured in the Flight-from-Feeder test, and WER was the trait measured in the Social Separation Test. Chrom = cattle chromosome.

Chrom	Traits Schmutz *et al.* (2001)	QTL Location (cM)	Traits present study	QTL Location (cM)
1	temperament + habituation	14	-	-
5	temperament + habituation	29	not tested	-
9	temperament + habituation	44	-	-
11	temperament + habituation	57	not tested	-
14	temperament only	19	WER	12
	temperament only	35	FF	58
15	temperament + habituation	12	FF	13

As mentioned above, QTLs for the temperament and habituation traits were also identified on chromosomes 1, 5, 9 and 11 in the Canadian herd (Schmutz *et al.*, 2001). Chromosomes 1 and 9 were also examined in this study, but no QTLs were found for FF or WER. There are a number of possible reasons for this. The behaviour measure of temperament taken in the Canadian herd is likely to reflect a mixture of a number of fear-related traits, such as fear of separation, fear of humans, fear of novelty (during the first test at least) and a reaction to physical confinement. In the present study, the tests used and behaviours measured were each designed to be measures of single traits, with reactions to other stimuli being minimised or controlled as far as possible. It is possible that the QTLs identified in the Canadian study each influenced different individual traits which were all reflected in the temperament measure used. The present study may have picked out only some of the QTLs that were identified in the Canadian herd because, even though the QTLs found in the Canadian Herd each influenced different traits, all contributed to, and were reflected in to some degree, the broad temperament measure used, which captured the animals behavioural responses to several temperament traits at the same time. The present analysis picks out different QTLs for sociality and fearfulness of humans, but (so far) none that influence both. These are considered to be separate traits and therefore are likely to be under the control of different genes.

This scenario had been found in mice. Turri *et al*. (2001b) carried out a genome scan for QTLs affecting behaviour shown in five tests of anxiety in mice. They found that some of the QTLs had narrow ranges of influence, and were specific to certain tests, whereas some influenced more than one test. One QTL influenced all the measures.

However, there are other reasons why QTLs identified by the Canadian study were not picked up here. As the genotyping data set was not complete, and the number of animals analysed was low, more QTLs may be detected in this study when the remaining genotype and pedigree errors are corrected. As different breed crosses were used in the two experiments, the same QTLs may not be segregating in both of the populations. It is also possible that some of the QTLs identified in the Canadian study or in this one are false positives.

The chromosomal locations of the QTLs identified in this study can also be compared to homologous regions of mouse chromosomes. QTLs for fear-related traits have been mapped to areas of mouse chromosomes 1, 12 and 15 in a number of independent studies (Flint *et al*., 1995; Gershenfeld *et al*., 1997; Gershenfeld & Paul, 1997; Talbot *et al*., 1999; Turri *et al*., 1999; 2001a; 2001b; see Chapter 1.7.3.3). The two QTLs identified on chromosome 14 in this study lie in a region which is homologous to an area on mouse chromosome 15, where a QTL for open-field behaviour has been mapped in mice (Turri *et al*., 1999). This provides further support for a QTL of interest at this location in this study.

The QTL results reported here, although limited, already show some interesting results, and hopefully will provide further impetus for similar studies to be carried out in different breed crosses.

6.5. Conclusions

The aim of this chapter was to investigate the genetic basis of temperament traits in cattle, by carrying out a QTL genome scan. Behavioural data collected throughout the three years of the project from the FF and SS Tests were used. Analysis prior to the QTL analysis showed that the behavioural data demonstrated suitable variation between animals and repeatability within animals and were therefore useful phenotypic trait measures.

Analysis of possible fixed effects to be included in the QTL analysis showed that no environmental factors (due to the way the animals were housed or tested) had a significant impact on the test scores. They also showed that factors intrinsic to the animals, such as their age or sire, had no impact on the test scores.

The QTL analysis was limited to four chromosomes with only incomplete genotyping information for the animals available. Despite this, three putative QTL locations were identified on two of the chromosomes. All three were in similar locations to QTLs for temperament traits in cattle identified in a previous study, with one lying extremely close. One of the QTLs lay in an area of a chromosome homologous to an area of a mouse chromosome where a QTL for fear-related behaviour has been fine-mapped. Therefore the putative QTL locations already identified have support from independent studies.

Results from further analyses, to be carried out when the genotyping dataset is complete, are keenly anticipated.

Chapter Seven: Discussion

It has long been assumed that cattle, as other animals, show individual differences in temperament traits. A huge number of studies have taken measures of temperament using a wide range of different test situations. However, many studies fail to either test for repeatability of the test measures, or give any details on whether the measures have previously been examined for consistency of response (for example, Boissy & Bouissou, 1995; Boivin *et al.*, 1992a; Boivin *et al.*, 1994; Kabuga & Appiah, 1992; Le Neindre *et al.*, 1995; Murphey *et al.*, 1981; Matthews *et al.*, 1997; Plusquellec & Bouissou, 2001). The study reported in this thesis has confirmed the existence of repeatable, independent behavioural traits in cattle that are consistent over long periods of time.

As behavioural responses are so easily affected by a range of factors other than the reactivity of the animal (such as it's actual psychobiological state at the time of testing, and the conditions of the particular test), it is essential to check whether responses being measured as traits meet the criteria of repeatability. The importance of this is increasingly being recognised, as demonstrated by the recent published studies mentioned in Chapter Three. The increasing recognition of standardised behavioural testing as a useful measurement tool was reflected by the 37th International Congress of the International Society for Applied Ethology held this year. A session was devoted to behavioural tests, and several papers were presented concerning repeatability and interpretation of test results, and the use of measures taken at an early age to predict future behaviour (Burman & Mendl, 2003; Koene *et al.*, 2003; Miller *et al.*, 2003; McBride & Wolf, 2003; Rooney *et al.*, 2003; van Reenen *et al.*, 2003).

However, the present study did not successfully validate the behavioural tests used. Cross-validation of the temperament measures being reflected was attempted, by looking for relationships between measures taken in different tests that were hypothesized to test the same trait. Only a weak relationship was found between the two tests thought to measure fearfulness of humans, while no relationship was found between measures from the two tests thought to measure sociality. It was hypothesized that the lack of relationships seen was due to the influence of additional factors, which differed between each test situation. It is very difficult to remove factors other than the one of interest from a test. For example, during most of the tests carried out in this thesis, two people were present to move the cattle in and out of pens, and hence fearfulness of humans is likely to play a part in the animals' responses, even in the tests attempting to measure sociality. In animals such as mice, it may be possible to set up purpose-built test environments, such as those with remotely-operated doors, to avoid human presence during testing. To build such apparatus for animals the size of cattle would be both prohibitively expensive and difficult to design.

An added complication concerning validation is that the mixture of traits measured by a test can vary depending on the past experiences of the animals. For example, while carrying out a sociality test, the presence of humans to move the animals in and out of the test pen may not affect the test responses of animals that have been handled frequently, but may have a large impact on the test responses of animals that have come into contact with humans very infrequently. This may explain the small relationship seen between the bulls' scores in the FF and SS Tests described in Chapter Four. This implies that tests not only need to be validated, but should be validated independently on each group of animals that they are used on.

If further time and access to these animals were available, carrying out a larger number of different tests may provide the cross-validation required. The Flight-from-Feeder (FF) Test and Handling (HA) Test may be validated using approach tests, (where the animal is allowed to approach a stationary person), or crush tests (where the animal is physically restrained in a crush cage with a person present). Pilot studies with a range of large bright objects may identify objects which elicit neophobia in these animals. A successful novel object test could perhaps then be validated by testing the animals with a different object in a another familiar environment, perhaps as an obstacle that the animals have to walk past in a familiar race (Jennings, 1999). In another attempt to validate the Social Separation (SS) Test, the Sociality (SO) Test could be carried out again, this time after allowing the animals to fully explore the corridor where the test was to be carried out, so the setting is as familiar to them as their home pen.

An effort should also be made to keep tests as simple and as quick to carry out as possible, so they are more likely to be used on other farms or in other experiments. The test set-ups used here were all very specific to the pens that the animals were housed in, but each could be easily adapted to be carried out on other farms. The most easily used is the FF Test. Flight tests are quick, and if adequately validated as a test of fearfulness of humans, could be incorporated into farm assessments, such as those carried out for quality assurance schemes, as one indicator of animal welfare. High levels of fearfulness of humans can be indicative of negative handling, and may limit productivity and welfare (see Chapter 1.4).

Two of the traits measured in Chapter Four, the FF and SS Scores, were shown to remain stable as the animals got older. However, the FF Score and the HA Latency could not be used as predictors of behaviour in the dairy parlour when being milked. This again goes back to the issue of validation. The 'Step, Kick' (SK) parlour responses did not correlate with FF Scores taken at the same age, and therefore clearly (in hindsight) were not going to be predicted by FF Scores taken at a younger age. In future, studies of prediction should be carried out using measures that have been previously validated at the same age already.

The QTL scan of four chromosomes revealed three putative QTL locations. Although they all had a low level of significance, they lie in the same chromosomal regions as QTLs identified in a previous study of cattle behaviour. Two lie in a region homologous to that of a mouse chromosomal region where QTLs for fearfulness traits have been repeatedly mapped. The detection of QTLs affecting theses behaviours is the first step towards identifying genes that affect temperament traits. Once the whole genome scan has been completed, there are a number of ways the information can be used. These include interpretation of behavioural functions, use in marker-assisted selection (MAS) programmes, and the identification of causative genes.

The QTLs found may be able to be used to validate behaviours seen in the different test procedures in retrospect, following an approach taken by Turri *et al.* (2001b). They examined whether behaviours in five tests of fear-related traits in mice were under the influence of a common set of genes. Five commonly-used tests of anxiety were used to phenotype the mice; the open-field arena, the elevated plus-maze, the square maze, the light dark box and the mirror chamber, and additionally control measures of behaviour in the home cage were taken. They identified significant QTLs on eight chromosomes. Some of these QTLs were test specific, most influenced more than one test, and one influenced all the measures. By examining the factors that were thought to be measured in each test, and matching which patterns of behaviour were influenced by which QTLs, they were able to interpret the functions of the QTLs. The authors interpreted different QTLs as being consistent with roles in visually mediated behaviour, locomotor activity, avoidance behaviour, exploratory behaviour, and also loci whose roles were more complex. Flint (2003) points out that drawing inferences about the existence of a psychological process [such as a fearfulness trait] from correlated behavioural responses lacks any way of identifying the biology common to that state. He argues that the assumption that different behaviours can represent manifestations of the same brain state can be validated by showing that the same QTLs affect a multitude of phenotypes in a theoretically predictable fashion.

Information about QTL locations is of direct use in applying to breeding programmes, using MAS. Several cattle and pig breeding companies are now using MAS to complement phenotypic selection of breeding animals (Andersson, 2001). The use of marker information makes it possible to select animals by their genotypes, without needing to take measures of phenotypes. Most applications of MAS in livestock are for traits such as those recorded here, that have moderate heritability and are difficult to improve by conventional means because they are difficult to record (Dekkers & Hospital, 2002). Using MAS in selection has been applied to traits such as muscling, growth, reproduction, meat and milk quality over the last few years (Dentine, 1999). Benefits have been highest when it is applied at a stage where normally no selection or random selection is applied, for example, young bulls entering a progeny test (Haley & Visscher, 1998). Previously, selection at this stage relied entirely on pedigree information on estimated breeding values, and so the choice between any number of full-sib young bulls was completely random. The use of marker information makes it possible to distinguish among these brothers and to select the most promising ones (Haley & Visscher, 1998). For use in MAS, markers need to be mapped to within 5 cM intervals, which would require fine-mapping techniques to be carried out on the regions of these putative QTLs.

When the whole genome scan has been completed, the ultimate goal is to eventually identify the causative genes affecting the traits, to allow investigation of the molecular mechanisms by which the gene exerts its effects. The poor precision of mapping studies such as this means that the molecular identification of these loci will rely on the exploitation of comparative data from mice or from humans (Andersson, 2001). Synteny homology between chromosomes of different species allows QTLs found in mouse models of fear traits to be used as candidate QTLs for fear in cattle. At the present time, no genes for behavioural traits have been mapped in mice. Since Flint *et al.* (1995) first identified QTLs for emotionality traits in mice, fine-mapping techniques have been used to refine the locations of those QTLs to intervals as small as 0.8 cM (Talbot *et al.*, 1999; Turri *et al.*, 1999). An interval of this size is small enough to consider possible candidate genes, and the identification of the relevant genes is expected in the near future (Mormède *et al.*, 2002).

An exciting advance in the search for genes affecting emotionality traits was the replicated association found between a functional polymorphism for the dopamine D4 receptor gene (*DRD4*) and novelty-seeking behaviour in humans (Benjamin *et al.*, 1996; Ebstein *et al.*, 1996). Another major finding has been the association between a functional polymorphism in a regulatory sequence for the seratonin transporter gene (*5-HTT*) and the emotional triad of neuroticism, depression and anxiety in humans (Lesch *et al.*, 1996). These genes can be used as candidate genes to guide investigations in other species.

Temperament traits such as those measured in this study are common to farm animals in virtually all husbandry systems, and the selection for these types of traits holds great potential for the improvement of farm animal welfare (Mills *et al.*, 1997). All husbandry systems expose animals to frightening stimuli, such as contact with humans or machinery. Other candidate temperament traits for QTL mapping and MAS may include aggression and maternal behaviour. Dairy companies are currently beginning to broaden their breeding indices to include measures that have relevance to welfare (Lawrence *et al.*, 2003). The putative QTLs identified in this study form a small but important step towards that goal. The difficulties of interpretation and validation of behaviours seen in behavioural tests may also be solved in retrospect by QTL analysis.

References

Albright, J.L. (1993). Dairy Cattle Husbandry. In: *Livestock Handling and Transport* (ed. T. Grandin). CAB International, Wallingford, Oxon, UK. p. 47.

Albright, J.L. & Arave, C.W. (1997). *The behaviour of cattle*. CAB International, Wallingford, Oxon, UK. 306 pp.

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2**: 130-138.

Andrade, O., Orihuela, A., Solano, J. & Galina, C.S. (2001). Some effects of repeated handling and the use of a mask on stress responses in zebu cattle during restraint. *Applied Animal Behaviour Science* **71**: 175-181.

Archer, J. (1973). Tests for emotionality in rats and mice: a review. *Animal Behaviour* **21**: 205-235.

Ashwell, M.S. & Van Tassell, C.P. (1999). Detection of putative loci affecting milk, health and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. *Journal of Dairy Science* **82**: 2497-2502.

Ashwell, M.S., Van Tassell, C.P. & Sonstegard, T.S. (2001). A genome scan to identify quantitative trait loci affecting economically important traits in a US Holstein population. *Journal of Dairy Science* **84**: 2535-2542.

Barendse, W. & Fries, R. (1999). Genetic linkage mapping, the gene maps of cattle and the lists of loci. In: *The Genetics of Cattle* (eds. R. Fries & A. Ruvinsky). CAB International, Wallingford, Oxon, UK. p. 329-364.

Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Ryan, A.M., Clayton, D., Li, L., Neibergs, H.L., Zhang, N., Grosse, W.M., Weiss, J., Creighton, P., McCarthy, F., Ron, M., Teale, A.J., Fries, R., McGraw, R.A., Moore, S.S., Georges, M., Soller, M., Womack, J.E. & Hetzel, D.J.S. (1994). A genetic linkage map of the bovine genome. *Nature Genetics* **6**: 227-235.

Barendse, W., Armitage, S.M., Ryan, A.M., Moore, S.M., Clayton, D., Georges, M., Womack, J.E. & Hetzel, J. (1993). A genetic map of DNA loci on bovine chromosome 1. *Genomics* **18**: 602-608.

Barendse, W., Vaiman, D., Kemp, S., Sugimoto, Y., Armitage, S., Williams, J., Sun, H., Eggen, A., Agaba, M., Aleyasin, A., Band, M., Bishop, M., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B., Drinkwater, R., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harlizius, B., Hawkins, G., Hetzel, J., Hirano, T., Hulme, D., Joergensen, C., Kessler, M., Kirkpatrick, B., Konfortov, B., Kostia, S., Kuhn, C., Lenstra, J., Leveziel, H., Lewin, H., Leyhe, B., Li, L., Martin Buriel, I., McGraw, R., Miller, R., Moody, D., Moore, S., Nakane, S., Nijman, I., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Soller, M., Teale, A., Thieven, I., Urquhart, B., Vage, D-I., Van de Weghe, A., Varvio, S., Velmalla, R., Vilkki, J., Weikard, R., Woodside, C., Womack, J., Zanotti, M., Zaragoza, P. (1997). A medium density genetic linkage map of the bovine genome. *Mammalian Genome* **8**: 21-28.

Benjamin, J., Li, L., Patterson, C., Greenburg, B.D., Murphey, D.L. & Hamer, D.H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics* **12**: 81-84.

Bishop, M.D., Kappes, S.M., Keele, J.W., Stone, R.T., Sunden, S.L.F., Hawkins, G.A., Toldo, S.S., Fries, R., Grosz, M.D., Yoo, J & Beattie, C.W. (1994). A genetic linkage map for cattle. *Genetics* **136**: 619-639.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y. & Leveziel, H. (2003). Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35**: 77-101.

Boissy, A. & Bouissou, M.-F. (1988). Effects of early handling on heifer's subsequent reactivity to humans and to unfamiliar situations. *Applied Animal Behaviour Science* **20**: 259-273.

Boissy, A. & Bouissou, M.F. (1994). Effects of androgen treatment on behavioural and physiological responses of heifers to fear-eliciting situations. *Hormones and Behaviour* **28**: 66-83.

Boissy, A. & Bouissou, M.-F. (1995). Assessment of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations. *Applied Animal Behaviour Science* **46**: 17-31.

Boissy, A. & Le Neindre, P. (1990). Social influences on the reactivity of heifers: implications for learning abilities in operant conditioning. *Applied Animal Behaviour Science* **25**: 149-165.

Boissy, A. & Le Neindre, P. (1997). Behavioural, cardiac and cortisol responses to brief peer separation and reunion in cattle. *Physiology & Behavior* **61**: 693-699.

Boissy, A. (1995). Fear and fearfulness in animals. *The Quarterly Review of Biology* **70**: 165-191.

Boissy, A. (1998). Fear and fearfulness in determining behavior. In: *Genetics and the Behavior of Domestic Animals* (ed. T. Grandin). Academic Press, San Diego, California. p. 67-111.

Boissy, A., Fisher, A., Bouix, J., Boivin, X. & Le Neindre, P. (2002). Genetics of fear and fearfulness in domestic herbivores. In: *Proceedings of the 7th World*

Congress on Genetics Applied to Livestock Production, Montpellier, France. August 19-23, 2002.

Boissy, A., Terlouw, C. & Le Neindre, P. (1998). Presence of cues from stressed conspecifics increases reactivity to aversive events in cattle: evidence for the existence of alarm substances in urine. *Physiology & Behaviour* **63**: 489-495.

Boivin, X., Le Neindre, P. Chupin, J.M., Garel, J.P. & Trillat, G. (1992a). Influence of breed and early management on ease of handling and open-field behaviour on cattle. *Applied Animal Behaviour Science* **32**: 313-323.

Boivin, X., Le Neindre, P. & Chupin, J.M. (1992b). Establishment of cattle-human relationships. *Applied Animal Behaviour Science* **32**: 325-335.

Boivin, X., Le Neindre, P., Garel, J.P. & Chupin, J.M. (1994). Influence of breed and rearing management on cattle reactions during human handling. *Applied Animal Behaviour Science* **39**: 115-122.

Breuer, K., Hemsworth, P.H., Barnett, J.L., Matthews, L.R. & Coleman, G.J. (2000). Behavioural response to humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science* **66**: 273-288.

Buchenauer, D. (1999). Genetics of behaviour in cattle. In: *The Genetics of Cattle* (eds. R. Fries & A. Ruvinsky). CAB International, Wallingford, Oxon, UK. p. 365-390.

Burman, O.H.P. & Mendl, M.T. (2003). Can pre-experimental social experience affect the reliability of behavioural tests? (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 72.

Burrow, H.M. (1997). Measurements of temperament and their relationships with performance traits of beef cattle. *Animal Breeding Abstracts* **65**: 477-495.

Casas, E., Shackleford, S.D., Keele, J.W., Stone, R.T., Kappes, S.M. & Koohmaraie, M. (2000). Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *Journal of Animal Science* **78**: 560-569.

Casas, E., Stone, R.T., Keele, J.W., Shackleford, S.D., Kappes, S.M. & Koohmaraie, M. (2001). A comprehensive search for quantitative trait loci affecting growth and carcass composition of cattle segregating alternative forms of the myostatin gene. *Journal of Animal Science* **79**: 854-860.

Churchill, G.A. & Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963-971.

de Groot, R. (2002). *Genetic dissection of anxiety related behaviour in mice*. Masters Thesis: Vrije Universiteit, Amsterdam, Netherlands. 45 pp.

de Groot, R., Sluyter, F., Lakenberg, N., Slagboom, E. & Boomsma, D. (2001). A review of QTL and knock-out studies for emotionality in mice (abs). *Behavior Genetics* **31**: 451-451.

de Passillé, A.M., Rushen, J. & Martin, F. (1995). Interpreting the behaviour of calves in an open-field test: a factor analysis. *Applied Animal Behaviour Science* **45**: 201-213.

DeFries, J.C., Gervais, M.C. & Thomas, E.A. (1978). Response to 30 generations of selection for open-field activity in laboratory mice. *Behaviour Genetics* **8**: 3-13.

Dekkers, J.C.M. & Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**: 22-32.

Dellmeier, G., Friend, T. & Gbur, E. (1990). Effects of changing housing on open-field behaviour of calves. *Applied Animal Behaviour Science* **26**: 215-230.

Dennison, S.G.C. (1986). *The development of behaviour patterns and an assessment of temperament of dairy heifers*. PhD Thesis: University of Edinburgh.

Dentine, M.R. (1999). Marker-assisted selection. In: *The Genetics of Cattle* (eds. R. Fries & A. Ruvinsky). CAB International, Wallingford, Oxon, UK. p. 497-510.

Ebstein, R.P., Novick, O., Umansky, R., Priel, B., Osher, Y., Blaine, D., Bennett, E.R., Nemanov, L., Katz, M. & Belmaker, R.H. (1996). Dopamine D4 receptor (*DRD4*) exon 111 polymorphism associated with the human personality trait novelty-seeking. *Nature Genetics* **12**: 78-80.

Elo, K., Vilkki, J., de Koning, D.J., Velmala, R.J. & Maki-Tanila, A.V. (1999). A quantitative trait locus for live weight maps to bovine chromosome 23. *Mammalian Genome* **10**: 831-835.

Erhard, H.W. & Schouten, W.G.P. (2001). Individual differences and personality. In: *Social Behaviour in Farm Animals* (ed. L.J. Keeling and H.W. Gonyou). CAB International, Wallingford, UK. 406 pp.

Falconer, D.S. (1989). *Introduction to Quantitative Genetics* (3rd edition). Longman Scientific and Technical, Essex, UK. 438 pp.

Feaver, J., Mendl, M., & Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour* **34**: 1016-1025.

Fisher, A.D., Morris, C.A. & Matthews, L.R. (2000). Cattle behaviour: comparison of measures of temperament in beef cattle. *Proceedings of the New Zealand Society of Animal Production* **60**: 214-217.

Fisher, A.D., Morris, C.A. & Matthews, L.R., Pitchford W.S. & Bottema, C.D.K. (2001). Handling and stress response traits in cattle: identification of putative genetic markers (abs). In: *Proceedings of the 35th International Congress of the ISAE* (eds. J.P. Garner, J.A. Mench & S.P. Heekin). University of California, Davis. August 4-9, 2001. p. 100.

Flint, J. (2003). Analysis of quantitative trait loci the influence animal behavior. *Journal of Neurobiology* **54**: 46-77.

Flint, J., Corley, R., DeFries, J.C., Fulker, D.W., Gray, J.A., Miller, S. & Collins, A.C. (1995). A simple genetic basis for a complex psychological trait in laboratory mice. *Science* **269**: 1432-1435.

Fordyce, G., Goddard, M.E. & Seifert, G.W. (1982). The measurement of temperament in cattle and the effect of experience and genotype. *Proceedings of the Australian Society of Animal Production* **14**: 329-332.

Forkman, B., Furuhaug, I.L. & Jensen, P. (1995). Personality, coping patterns and aggression in piglets. *Applied Animal Behaviour Science* **45**: 31-42.

Fowler, J. & Cohen, L. (1990). *Practical Statistics for Field Biology*. Wiley, Chichester, UK. 226 pp.

Frey, D.F. & Pimentel, R.A. (1978). *Principal Component Analysis and Factor Analysis*. In: P.W. Colgan (ed.), Quantitative Ethology. Wiley, New York, p. 219-245.

GenStat 5 Committee (1993). *GenStat 5 Release 3 Reference Manual*. Clarendon Press, Oxford, UK. 796 pages.

Georges, M. (1998). Mapping genes underlying production traits in livestock. In: *Animal Breeding: Technology for the 21st Century* (ed. A.J. Clark). Harwood Academic Publishers, Amsterdam, The Netherlands.

Georges, M. (2000). Recent progress in livestock genomics and potential impact on breeding programs. *Theriogenology* **55**: 15-21.

Georges, M., Neilsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., Womack, J.E. & Hoeschele, I. (1995). Mapping quantitative trait loci controlling milk production by exploiting progeny testing. *Genetics* **4**: 206-210.

Gershenfeld, H.K., & Paul, S.M. (1997). Mapping Quantitative Trait Loci for fear-like behaviours in mice. *Genomics* **46**: 1-8.

Gershenfeld, H.K., Neumann, P.E., Mathis, C., Crawley, J.N., Li, X. & Paul, S.M. (1997). Mapping Quantitative Trait Loci for open-field behaviour in mice. *Behaviour Genetics* **27**: 201-210.

Goldsmith, H.H. (1989). Behaviour-genetic approaches to temperament. In: *Temperament in Childhood.* (ed. G.A. Kohnstamm, J.E. Bates & M.K. Rothbart). Wiley, New York, USA. p. 111-132.

Goldsmith, H.H., Buss, A.H., Plomin, R., Rothbart, M.K., Thomas, A., Chess, S., Hinde, R.A. & McCall, R.B. (1987). Roundtable: what is temperament? *Child Development* **58**: 505-529.

Gosling, S.D. (2001). From mice to men: what can we learn about personality from animal research? *Psychological Bulletin* **127**: 45-86.

Green, P., Falls, K. & Crooks, S. (1990). Cri-map, Version 2.4. Washington University School of Medicine, St Louis, MO.

Grignard, L., Boissy, A., Boivin, X., Garel, J.P. & Le Neindre, P. (2000). The social environment influences the behavioural responses of beef cattle to handling. *Applied Animal Behaviour Science* **68**: 1-12.

Grignard, L., Boivin, X., Boissy, A. & Le Neindre, P. (2001). Do beef cattle react consistently to different handling situations? *Applied Animal Behaviour Science* **71**: 263-276.

Grupe, S., Panicke, L., Dietl, G., Kuhn, C., Gulard, V., Schwerin, M. (1998). Identification of loci with significant effects on stillbirth and calving difficulties in Holstein cattle. *Archiv Fur Tierzucht - Archives Of Animal Breeding* **41**: 151-158.

Haley, C.S. & Andersson, L. (1997). Linkage Mapping of quantitative trait loci in plants and animals. In: *Genome Mapping: A Practical Approach* (ed. P.H. Dear). Oxford University Press Inc., New York. p. 49-71.

Haley, C.S., Knott, S.A. & Elsen, J-M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**: 1195-1207.

Haley, C.S. & Visscher, P.M. (1998). Strategies to utilise marker-quantitative trait loci associations. *Journal of Dairy Science* **81**: 85-97.

Hansen, S.W. (1996). Selection for behavioural traits in farm mink. *Applied Animal Behaviour Science* **49**: 137-148.

Hemsworth, P.H., Barnett, J.L., Tilbrook, A.J. & Hansen, C. (1989). The effects of handling by humans at calving and during milking on the behaviour and milk cortisol concentrations of primiparous dairy cows. *Applied Animal Behaviour Science* **22**: 313-326.

Hemsworth, P.H. & Coleman, G.J. (1998). *Human animal interactions: the stockperson and the productivity and welfare of intensively farmed animals.* CAB International, Wallingford, UK. 152 pp.

Hemsworth, P.H., Coleman, G.J., Barnett, J.L. & Borg, S. (2000). Relationships between human-animal interactions and productivity of commercial dairy cows. *Journal of Animal Science* **78**: 2821-2831.

Hemsworth, P.H., Coleman, G.J., Barnett, J.L., Borg, S. & Dowling, S. (2002). The effects of cognitive behavioural intervention on the attitude and behaviour of stockpersons and the behaviour and productivity of commercial dairy cows. *Journal of Animal Science* **80**: 68-78.

Hemsworth, P.H., Hansen, C. & Barnett, J.L. (1987). The Effects of Human Presence at the Time of Calving of Primiparous Cows on their Subsequent Behavioural Response to Milking. *Applied Animal Behaviour Science* **18**: 247-255.

Hemsworth, P.H., Price, E.O. & Borgwardt, R. (1996). Behavioural responses of domestic pigs and cattle to humans and novel stimuli. *Applied Animal Behaviour Science* **50**: 43-56.

Heyen, D.W., Weller, J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da, Y., Wiggans, G.R., Van Raden, P.M. & Lewin, H.A. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics* 1: 165-175.

Hocking, P.M., Channing, C.E., Waddington, D. & Jones, R.B. (2001). Age-related changes in fear, sociality and pecking behaviours in two strains of laying hen. *British Poultry Science* **42**: 414-423.

Hopster, H. & Blokhuis, H.J. (1993). Consistent stress responses of individual dairy cows to social isolation. In: *Proceedings of the International Congress on Applied Ethology* (eds. M. Nichelmann, H.K. Wieranga & S. Braun). Berlin, Germany. p. 123-126.

Jennings, M. (1999). *Individual variation in maternal behaviour in mammals, with a special focus on maternal temperament in cattle*. MSc Thesis, University of Edinburgh. 50 pp.

Jensen, M.B. (2001). A note on the effect of isolation during testing and length of previous confinement on locomotor behaviour during open-field test in dairy calves. *Applied Animal Behaviour Science* **70**: 309-315.

Johnston, T.E., Defries, J.C. & Markel, P.D. (1992). Mapping quantitative trait loci for behavioural traits in the mouse. *Behavior Genetics* **22**: 635-653.

Jones, R.B. (1987). The assessment of fear in adult laying hens: correctional analysis of the methods and measures. *British Poultry Science* **28**: 319-326.

Jones, R.B. (1988). Repeatability of fear ranks among adult laying hens. *Applied Animal Behaviour Science* **19**: 297-304.

Jones, R.B. (1996). Fear and adaptability in poultry: insights, implications and imperatives. *World's Poultry Science Journal* **52**: 131-174.

Jones, R.B., Marin, R.H., Satterlee, D.G. & Cadd, G.G. (2002). Sociality in Japanese quail (*Coturnix japonica*) genetically selected for contrasting adrenocortical responsiveness. *Applied Animal Behaviour Science* **75**: 337-346.

Jones, R.B. & Waddington, D. (1992). Modification of fear in domestic chicks, *Gallus gallus domesticus*, via regular handling and early environmental enrichment. *Animal Behaviour* **43**: 1021-1033.

Kabuga, J.D. & Appiah, P. (1992). A note on the ease of handling and flight distance of *Bos indicus*, *Bos taurus* and their crossbreds. *Animal Production* **54**: 309-311.

Kappes, S.M., Keele, J.W., Stone, R.T., McGraw, R.A., Sonstegard, T.S., Smith, T.P.L., Lopez-Corrales, N.L., Beattie, C.W. (1997). A second-generation linkage map of the bovine genome. *Genome Research* **7**: 235-249.

Keele, J.W., Shackleford, S.D., Kappes, S.M., Koohmaraie, M. & Stone, R.T. (1999). A region on bovine chromosome 15 influences beef longissimus tenderness in steers. *Journal of Animal Science* **77**: 1364-1371.

Kiley-Worthington, M. & de la Plain, S. (1983). *The Behaviour of Beef Suckler Cattle (Bos Taurus).* Birkhäuser Verlag Basel, Switzerland. 195 pp.

Kilgour, R. (1975). The open-field test as an assessment of the temperament of dairy cows. *Animal Behaviour* **23**: 615-624.

Kleinbauen, D.G., Kupper, L.L. & Muller, K.E. (1988). *Applied Regression Analysis and Other Multivariate Methods*. PWS Kent Publishing Co., Boston, USA. 718 pp.

Klungland, H., Sabry, A., Heringstad, B., Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Odegard, J., Klemetsdal, G., Schulman, N., Vilkki, J., Ruane, J., Aasland, M., Ronningen, K. & Lien, S. (2001). Quantitative trait loci affecting clinical mastitis and somatic cell count in diary cattle. *Mammalian Genome* **12**: 837-842.

Koene, P., Bokkers, E.A.M., Bolhuis, J.E., Rodenburg, T.B., Urff, E.M. & Zimmerman, P.H. (2003). The open-field; test of emotionality or an approach-avoidance conflict? (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 86.

Krohn, C.C., Jago, J.G. & Boivin, X. (2001). The effect of early handling on the socialisation of young calves to humans. *Applied Animal Behaviour Science* **74**: 121-133.

Kuhn, C., Bennewitz, J., Reinsch, N., Xu, N., Thomson, H., Looft, C., Brockman, G.A., Schwerin, M., Weimann, C., Hiendieder, S., Erhardt, G., Medjugorac, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blumel, J., Kalm, E. (2003). Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *Journal of Dairy Science* **86**: 360-368.

Lander, E. & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**: 241-247.

Lander, E.S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185-199.

Lawrence, A.B., Simm, G. & Conington, J. (2003). Breeding and Animal Welfare: a review of options in farm animals (abs). In: *Proceedings of 'Science in the service of animal welfare',* UFAW symposium, Edinburgh, UK. 2-4 April, 2003.

Lawrence, A.B., Terlouw, E.M.C. & Illius, A.W. (1991). Individual differences in behavioural responses of pigs exposed to non-social and social challenge. *Applied Animal Behaviour Science* **30**: 73-86.

Le Neindre, P. (1989). Influence of rearing conditions and breed on social behaviour and activity of cattle in novel environments. *Applied Animal Behaviour Science* **23**: 129-140.

Le Neindre, P., Trillat, G., Sapa, J., Ménissier, F., Bonnett, J.N. & Chupin, J.M. (1995). Individual differences in docility in Limousin cattle. *Journal of Animal Science* **73**: 2249-2253.

Lesch, K.P., Bengel, D., Heils, A., Zhang Sabol, S., Greenburg, B.D., Petri, S., Benjamin, J., Muller-Reible, C.R., Hamer, D.H. & Murphy, D.L. (1996). A gene regulatory region polymorphism alters seratonin transporter expression and is associated with anxiety-related personality traits. *Science* **274**: 1527-1530.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., MA, USA. 980 pp.

Lyons, D.M. (1989). Individual differences in temperament of dairy goats and the inhibition of milk ejection. *Applied Animal Behaviour Science* **22**: 269-282.

Lyons, D.M., Price, E.O. & Moberg, G.P. (1988). Individual differences in temperament of domestic dairy goats: constancy and change. *Animal Behaviour* **36**: 1323-1333.

MacNeil, M.D. & Grosz, M.D. (2002). Genome-wide scans for QTL affecting carcass traits in Hereford x composite double backcross populations. *Journal of Animals Science* **80**: 2316-2324.

Manly, B.F.J. (1986). *Multivariate Statistical Methods: a Primer*. Chapman & Hall, London, UK. 159 pp.

Manteca, X. & Deag, J.M. (1993). Individual differences in temperament of domestic animals: a review of the methodology. *Animal Welfare* **2**: 247-268.

Marin, R.H., Freytes, P., Guzman, D. & Jones, R.B. (2001). Effects of an acute stressor on fear and on the social reinstatement responses of domestic chicks to cagemates and strangers. *Applied Animal Behaviour Science* **71**: 57-66.

Martin, P. & Bateson, P. (1993). *Measuring Behaviour: An Introductory Guide* (2nd edition). Cambridge University Press, Cambridge, UK. 222 pp.

Matthews, L.R., Carragher, J.F. & Slater, J.L. (1997). Effects of flightiness, sociability and previous handling experience on the behaviour of cattle in yards (abs). In: *Proceedings of the 31st International Congress of the ISAE* (ed. P.H. Hemsworth, M. Spinka, & L. Kostal). Prague, Czech Republic, 13-16 August, 1997. p. 94.

McBride, S.D. & Wolf, B. (2003). Repeatability of factor analysis; a study of ovine temperament (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 95.

McCrae, R.R. & Costa, P.T., Jr. (1985). Updating Norman's 'adequate taxonomy': intelligence and personality dimensions in natural languages and in questionnaires. *Journal of Personality & Social Psychology* **49**: 710-421.

Mendl, M. & Harcourt, R. (1988). Individuality in the domestic cat. In: *The Domestic Cat: the biology of its behaviour* (eds. D.C. Turner & P. Bateson). Cambridge University Press. p. 41-54.

Milhaud, J.M., Halley, H. & Lassalle, J.M. (2002). Two QTLs located on chromosomes 1 and 5 modulate different aspects of the performance of mice of the B x D ty RI strain series in the Morris navigation task. *Behavior Genetics* **32**: 69-78.

Miller, K.A., Mench, J.A. & Garner, J.P. (2003). The test-retest reliability of behavioural tests: a critical evaluation based on tests of fearfulness in quail (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 76.

Mills, A.D., Beiharz, R.G. & Hocking, P.M. (1997). Genetic Selection. In: *Animal Welfare*. (eds. M.C. Appleby & B.O. Hughes). CAB International, Wallingford, Oxon, UK. p. 219 - 231.

Mills, A.D., Jones, R.B. & Faure, J.M. (1995). Species specificity of social reinstatement in Japanese quail *Coturnix japonica* genetically selected for high or low levels of social reinstatement behaviour. *Behavioural Processes* **34**: 13-22.

Mills, A.D., Jones, R.B., Faure, J.M. & Williams, J.B. (1993). Responses to isolation in Japanese quail genetically selected for high or low sociality. *Physiology & Behavior* **53**: 183-189.

Mormède, P., Courvoisier, H., Ramos, A., Marissal-Arvy, N., Ousova, O., Desautes, C., Duclos, M., Chaouloff, F. & Moisan, M.-P. (2002). Molecular genetic approaches to investigate individual variations in behavioural and neuroendocrine stress responses. *Psychoneuroendocrinology* **27**: 563-583.

Munksgaard, L. & Jensen, M.B. (1996). The use of "open field" tests in the assessment of welfare of cattle. *Acta Agriculturae Scandinavica Section A: Animal Sciences*. Supplement **27**: 82-85.

Munksgaard, L., de Passillé, A.M., Rushen, J., Herskin, M.S. & Kristensen, A.M. (2001). Dairy cows' fear of people: social learning, milk yield and behaviour at milking. *Applied Animal Behaviour Science* **73**: 15-26.

Murphey, R.M., Duarte, F.A.M. & Penedo, M.C.T. (1980). Approachability of bovine cattle in pastures: breed comparisons and a breed x treatment analysis. *Behavior Genetics* **10**: 171-181.

Murphey, R.M., Duarte, F.A.M. & Penedo, M.C.T. (1981). Responses of cattle to humans in open spaces: breed comparisons and approach-avoidance relationships. *Behavior Genetics* **11**: 37-48.

Newman, S. (1994). Quantitative- and molecular-genetic effects on animal well-being: adaptive mechanisms. *Journal of Animal Science* **72**: 1641-1653.

Nguyen, P.V. & Gerlai, R. (2002). Behavioural and physiological characterization of inbred mouse strains: prospects for elucidating the molecular mechanisms of mammalian learning and memory. *Genes Brain & Behavior* **1**: 72-81.

Nicholas, F.W. (2000). *Introduction to Veterinary Genetics*. Blackwell Science Ltd., Oxford, UK. 316 pp.

Nicoll, L.J. (1999). *Management procedures and standards for the Roslin Institute Genome Mapping Resource Herd*. Roslin Institute, internal document.

Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Klungland, H., Svendsen, M., Adnoy, T., Sabry, A., Klemetsdal, G., Schulman, N., Kramer, W., Thaller, G., Ronningen, K. & Lien, S. (2002). A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *Journal of Dairy Science* **85**: 3124-3130.

Patterson, H.D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545-554.

Pervin, L.A. (1996). *The Science of Personality*. John Wiley & Sons, Inc., USA. 471 pp.

Plomin, R., DeFries, J.C., McClearn, G.E. & Rutter, M. (1997). *Behavioural Genetics* (3rd ed.). W.H. Freeman and Company, New York, USA. 367 pp.

Plusquelle, P. & Bouissou, M-F. (2001). Behavioural characteristics of two dairy breeds of cows selected (Hérens) or not (Brune des Alpes) for fighting and dominancy ability. *Applied Animal Behaviour Science* **72**: 1-21.

Purcell, D., Arave, C.W. & Walters, J.L. (1988). Relationship of three measures of behaviour to milk production. *Applied Animal Behaviour Science* **21**: 307-313.

Queller, D.C., Strassmann, J.E. & Hughes, C.R. (1993). Microsatellites and Kinship. *Trends in Ecology & Evolution* **8**: 285-288.

Ramos, A. & Mormède, P. (1998). Stress and emotionality: a multidimensional and genetic approach. *Neuroscience and Biobehavioural Reviews* **22**: 33-57.

Redbo, I. (1998). Relations between oral stereotypies, open-field behaviour, and pituitary-adrenal system in growing dairy cattle. *Physiology & Behavior* **64**: 273-278.

Rooney, N.J., Gaines, S.A. & Bradshaw, J.W.S. (2003). How predictive are puppy tests?: evidence from a puppy walking programme for military search dogs (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 41.

Rushen, J. (2000). Some issues in the interpretation of behavioural responses to stress. In: *The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare* (ed. G.P. Moberg & J.A. Mench). CAB International, Wallingford, UK. p. 23-42.

Rushen, J., de Passillé, A.M. & Munksgaard, L. (1999). Fear of people by cows and effects on milk yield, behavior and heart rate at milking. *Journal of Dairy Science* **82**: 720-727.

Rushen, J., Munksgaard, L., Marnet, P.G. & de Passillé, A.M. (2001). Human contact and the effects of acute stress on cows at milking. *Applied Animal Behaviour Science* **73**: 1-14.

Sánchez, J.M., Castro, M.J., Alonso, M.E. & Gaudioso, V.R. (1996). Adaptive metabolic responses in females of the fighting breed submitted to different sequences of stress stimuli. *Physiology & Behavior* **60**: 1047-1052.

Schmutz, S.M., Stookey, J.M., Winkelman-Sim, D.C., Plante, Y. & Buchanan, F.C. (2000). A QTL for behaviour maps to cattle chromosome 9 (abs). In: *Proceedings of 27th International Conference on Animal Genetics 'Animal Genomics: Synthesis of Past, Present and future directions'.* University of Minnesota, USA. July 22-26th 2000. p. 49.

Schmutz, S.M., Stookey, J.M., Winkelman-Sim, D.C., Waltz, C.S., Plante, Y. & Buchanan, F.C. (2001). A QTL study of cattle behavioural traits in embryo transfer families. *Journal of Heredity* **92**: 290-292.

Schrooten, C., Bovenhuis, H., Coppieters, W. & Van Arendonk, J.A.M. (2000). Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *Journal of Dairy Science* **83**: 795-806.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M. & Visscher, P.M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* **18**: 339-340.

Simm, G. (1998). *Genetic improvement of cattle and sheep.* Farming Press, Ipswich, UK. 433 pp.

Stevenson-Hinde, J., Stillwell-Barnes, R. & Zunz, M (1980). Individual differences in young rhesus monkeys: consistency and change. *Primates* **21**: 498-509.

Stone, R.T., Keele, J.W., Shackleford, S.D., Kappes, S.M. & Koohmaraie, M. (1999). A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *Journal of Animal Science* **77**: 1379-1384.

Strelau, J. (1983). *Temperament, personality and arousal.* Academic Press, London, UK. 375 pp.

Talbot, C.J., Nicod, A., Cherny, S.S., Fulker, D.W., Collins, A.C. & Flint, J. (1999). High-resolution mapping of quantitative trait loci in outbred mice. *Nature Genetics* **21**: 305-308.

Turri, M.G., Henderson, N.D., DeFries, J. & Flint, J. (2001a). Quantitative trait locus mapping in laboratory mice derived from a replicated selection experiment for open-field activity. *Genetics* **158**: 1217-1226.

Turri, M.G., Datta, S.R., DeFries, J., Henderson, N.D. & Flint, J. (2001b). QTL analysis identifies multiple behavioural dimensions in ethological tests of anxiety in laboratory mice. *Current Biology* **11**: 725-734.

Turri, M.G., Talbot, C.J., Radcliffe, R.A., Wehner, J.M. & Flint, J. (1999). High-resolution mapping of quantitative trait loci for emotionality in selected strains of mice. *Mammalian Genome* **10**: 1098 - 1101.

Van Reenen, C.G., Van der Werf, J.T.N., Bruckmaier, R.M., Hopster, H., Engel, B., Noordhuizen, J.P.T.M. & Blokhuis, H.J. (2002). Individual differences in behavioural and physiological responsiveness of primiparous dairy cows to machine milking. *Journal of Dairy Science* **85**: 2551-2561.

van Reenen, C.G., van der Werf, J.T.N., Ruis-Heutinck, L.F.M., Jones, R.B. & Blokhuis, H.J. (2003). Responsiveness of heifer calves to behavioural tests may predict adult reactivity to machine milking (abs). In: *Proceedings of the 37th International Congress of the ISAE* (eds. V. Ferrante, E. Canali, C. Carenzi, S. Mattiello, M. Minero, C. Palestrini, M.V. Tosi & M. Verga). Abano Terme, Italy. 24-28 June, 2003. p. 64.

Vanderwert, W., Berger, L.L., McKeith, F.K., Baker, A.M., Gonyou, H.W. & Bechtel, P.J. (1985). Influence of Zeronol implants on growth, behaviour and carcass traits in Angus and Limousin bulls and steers. *Journal of Animal Science* **61**: 310-319.

Veissier, I. & Le Neindre, P. (1992). Reactivity of Aubrac heifers exposed to a novel environment alone or in groups of four. *Applied Animal Behaviour Science* **33**: 11-15.

Viitala, S.M., Schulman, N.F., de Koning, D.J., Elo, K., Kinos, R., Virta, A., Virta, J., Maki-Tanila, A. & Vilkki, J.H. (2003). Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *Journal of Dairy Science* **86**: 1828-1836.

Waiblinger, S., Menke, C. & Coleman, G. (2002). The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Applied Animal Behaviour Science* **79**: 195-219.

Weber, J.L. & May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

Wemelsfelder, F. & Birke, L. (1997). Environmental Challenge. In: *Animal Welfare* (eds. M.C. Appleby & B.O. Hughes). CAB International, Wallingford, UK. p. 35-47.

Willis, G.L. (1983). A possible relationship between the flinch, step and kick response and milk yield in lactating cows. *Applied Animal Ethology* **10**: 287-290.

Wilsson, E. & Sundgren, P. (1997). The use of a behaviour test for the selection of dogs for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Applied Animal Behaviour Science* **53**: 279-295.

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L.A., Grignola, F.E., Uimari, P., Thaller, G. & Bishop, M.D. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* **149**: 1959-1973.

Zuckerman, M. (1991). *Psychobiology of personality*. Cambridge University Press, Cambridge, UK. 482 pp.

Publications

The following publications have arisen from the work described in this thesis:

Ball, N., Haskell, M.J., Appleby, M.C. & Williams, J.L. (2001). Investigation of Temperament Traits in Cattle for Quantitative Trait Locus (QTL) Identification (abs). In: *Proceedings of the 35th International Congress of the ISAE*, University of California, Davis, USA. August 4-9th, 2001. p. 130.

Ball, N., Haskell, M.J., Williams, J.L. & Deag, J. (2002). Investigating Temperament Traits in Cattle for Quantitative Trait Loci (QTL) Identification (abs). In: *Proceedings of the British Society for Animal Science Annual Meeting*, York, UK. April 8-10th, 2002. p. 66.

Ball, N., Haskell, M.J., Deag, J. & Williams, J.L. (2002). Measuring Temperament Traits in Cattle for QTL identification. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France. August 19-23rd, 2002.

Investigation of temperament traits in cattle for quantitative trait locus (QTL) identification

N. Ball[1], M.J. Haskell[2], M.C. Appleby[3] & J.L. Williams[1]

[1] Genomics and Bioinformatics Department, Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, Scotland
[2] Animal Behaviour Sciences Department, SAC, Bush Estate, Midlothian, EH26 OPH, Scotland
[3] Institute of Ecology and Resource Management, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, Scotland

Address correspondence to Miss Nia Ball, e-mail: nia.ball@bbsrc.ac.uk

. . . .

The temperament of animals affects the way they react to environmental challenge. Behavioural responses of cattle to handling and management systems on farms are presumed to reflect underlying temperament traits such as fear. As some temperament measures have a strong genetic component, selection for desirable behavioural responses could increase the ability of animals to cope with stressors encountered on farms. The aim of this study is to obtain reliable measures of temperament traits in cattle, and identify the chromosomal location of the genes involved in these traits. Behavioural data for quantitative trait loci (QTL) analysis must be shown to reflect underlying traits, by demonstrating repeatability and validity of tests, and show variation within the population.

Temperament tests were carried out on 50 10-month-old F_2 bull calves from a Holstein x Charolais resource herd. Four tests were used; a flight-from-feeder test (FF), a social separation test (SS), a novel object test (NO) and a handling test (H). The variables measured included locomotory, escape and exploratory behaviours, which are presumed to reflect underlying traits of fear, aggression and exploratory motivation. Each test was validated by comparing responses between tests, and was repeated 3 times per individual to assess repeatability.

Each test showed a wide range of behavioural responses within the herd. Repeatability values were found to be high for all tests. For instance, high repeatability was shown in FF scores (REML: $r=0.524 \pm 0.080$; $N=48$), and in duration of locomotion in the SS (Kendall's coefficient of concordance: $W=0.740$; Chi-square=84.26; DF=38; $p<0.001$). Data on cross-validation of tests will also be presented.

The variability and repeatability of these behavioural variables show that they are suitable for potential identification of QTLs. These results, along with those from another 400 calves, will be correlated with the genotyping of 200 microsatellite markers to localise QTLs for these traits.

Investigating temperament traits in cattle for quantitative trait loci (QTL) identification

N. Ball[1], M.J. Haskell[2], J.L. Williams[1], J.M. Deag[3]
[1] *Genomics and Bioinformatics Department, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, Scotland*
[2] *Animal Behaviour Sciences Department, SAC, Bush Estate, Midlothian EH26 OPH, Scotland*
[3] *Institute of Cell, Animal and Population Biology, The University of Edinburgh, Ashworth Laboratories, King's Buildings, West Mains Road, Edinburgh EH9 3JT, Scotland*

Introduction Farm animals show individual variation in their behavioural responses to handling and management systems on farms. These behavioural responses are presumed to reflect underlying temperament traits such as fear or aggression. Information about the location of genes that influence temperament traits could be used in selective breeding programmes to improve animal welfare, as selection for desirable behavioural responses would increase the ability of animals to cope with stressors encountered on farms. The aims of this study were to obtain reliable temperament measurements in cattle using behavioural tests, and to use this data to localise the genes (quantitative trait loci) that are involved in such traits.

Behavioural data obtained in temperament tests must be shown to reflect underlying traits by demonstrating intra-animal repeatability, inter-animal variability and validity. The objectives of this experiment were i) to carry out four behaviour tests on a group of heifers, and examine the repeatability, variability and validity of the results obtained; ii) to correlate the behavioural data with genotyping data from a large number of heifers to look for associations between behavioural phenotypes and genetic markers. Associations localise quantitative trait loci (QTLs), or regions of the genome, that are involved in these traits.

Materials and Methods Temperament tests were carried out on 54 12-month-old heifer calves from the second generation cross of a Holstein x Charolais resource herd. Four tests were used; a Flight-from-Feeder test (FF), a Social Separation Test (SS), a Novel Object Test (NO) and a Handling Test (H). The variables measured included locomotory and escape behaviours, and are presumed to reflect underlying traits of fear, aggression and exploratory motivation. Each test was repeated twice per individual to assess repeatability. Restricted Maximum Likelihood (REML) was used to calculate repeatability values, and where appropriate Principle Components Analysis (PCA) was used to group behaviours of a common motivational background.

Results The FF Test demonstrated a wide range of scores between individuals (Fig. 1), that had high repeatability (REML; $r = 0.56 \pm 0.12$). The SS Test also showed a wide range of behavioural response with a high repeatability of durations of 'fearful' behaviours ($r = 0.63 \pm 0.11$; Fig. 2). The NO test showed low variability of response, and a low repeatability of 'duration of contact time' ($r = 0.26 \pm 0.12$). The HA Test showed a wide variability of response, and a low repeatability of 'latency to be held in corner of the pen' ($r = 0.36 \pm 0.15$). No correlations were found between the different test results, suggesting that different traits were being measured in each.

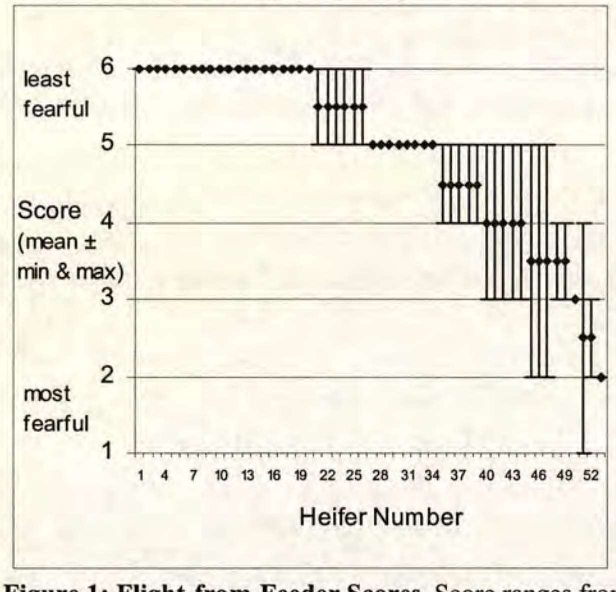


Figure 1: Flight-from-Feeder Scores. Score ranges from 1 (heifer moves away when the observer is >2m away) to 6 (heifer doesn't move back when observer touches on the head)

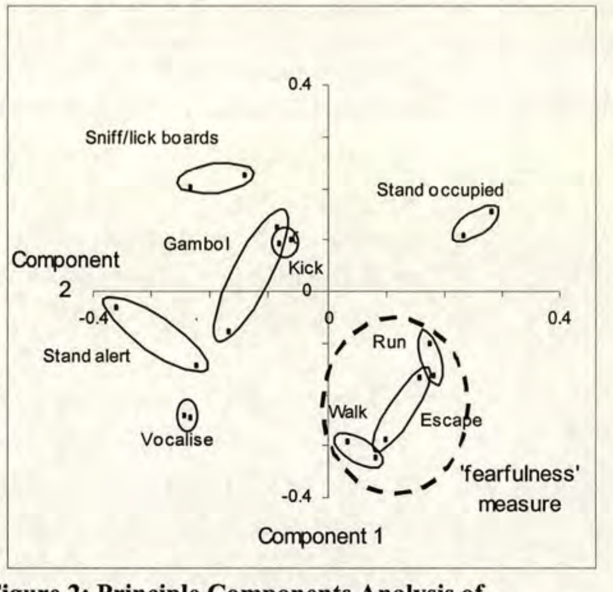


Figure 2: Principle Components Analysis of behaviours that occur in the Social Separation Test. Durations of behaviours shown in the 2 tests are ringed. Component 1 explains 26.4% of the variation, Component 2, 11.2%. The summed durations of Escape, Run and Walk are used as a measure of fearfulness (dashed circle).

Conclusion The variability and repeatability of the behavioural variables from the FF and SS Tests show that they are suitable for use in potential identification of QTLs. Further validation of the tests will also be presented, along with preliminary QTL analysis carried out using the data from these tests and genotyping data from 200 microsatellite markers.

MEASURING TEMPERAMENT TRAITS IN CATTLE FOR QTL IDENTIFICATION

N. Ball[1], M.J. Haskell[2], J.M. Deag[3] and J.L. Williams[1]

[1] Division of Genomics and Bioinformatics, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, Scotland, UK
[2] Animal Behaviour Sciences Dept., SAC, Bush Estate, Midlothian EH26 OPH, Scotland, UK
[3] Institute of Cell, Animal and Population Biology, The University of Edinburgh, Ashworth Laboratories, King's Buildings, West Mains Road, Edinburgh EH9 3JT, Scotland, UK

INTRODUCTION

Cattle show individual variation in their behavioural responses to handling and management systems on farms. These behavioural responses are presumed to reflect underlying temperament traits such as fear or aggression. Differences between breeds in behavioural traits demonstrate that variation in temperament has a genetic component, and heritabilities for such traits have generally been estimated as moderate to high (Burrows, 1997). Selection for favourable behavioural phenotypes would increase the ability of animals to cope with stressors encountered in modern agricultural systems, improving animal welfare and productivity, and human safety when handling stock. Hence information about the genetic loci that influence temperament may be of use in selective breeding programmes to select for animals with temperaments better suited to their environment.

Behavioural data obtained in temperament tests must be shown to reflect underlying traits by demonstrating intra-animal repeatability, inter-animal variability and validity. The aims of this study were to obtain reliable temperament measures in cattle using behavioural tests, and to use this data to localise quantitative trait loci (QTLs) involved in temperament. Two behavioural tests were carried out on groups of bull and heifer calves, and the variability and repeatability of the results were examined. The data from the behavioural phenotypes will be correlated with genetic markers to localise the QTLs involved.

MATERIALS AND METHODS

Behavioural responses to two important aspects of handling, human approach and separation from the group, were measured. A Flight-from-Feeder (FF) Test and a Social Separation (SS) Test were carried out on 49 bull calves and 56 heifer calves at approximately 11 months of age. The animals were second generation cross-bred Charolais x Holstein from a resource herd, and included F_2 calves and backcrosses to each parental breed. These animals were the first group of the 400 second-generation offspring in the herd. The bulls were reared with their dams at pasture for the first six months and then housed in groups. The heifers were reared on a dairy bucket system from 24 hours after birth and grouped at two weeks.

The FF Test. This test measured fearfulness of human approach, by measuring how quickly an animal moved away from a feeder when an observer approached. Each animal was given a score between 1 and 6 from a categorical scale, 1 being the score obtained when the animal moved back when the observer was >2m away, and 6 being the score obtained if the animal didn't move back when the observer touched it on the head.

The SS Test. This test measured how an animal responded to separation from its group when placed alone in its home pen for 5 minutes. Behaviours displayed during this time were continuously recorded, and total durations calculated. Principle Components Analysis (PCA) was used to group behaviours of a common motivational background.
Each test was carried out twice on each animal. Restricted Maximum Likelihood (REML) was used to calculate repeatability values.

RESULTS AND DISCUSSION

The FF Test demonstrated a wide range of scores between individuals (Figure 1a+b). High inter-animal repeatability of test scores was seen for both the bulls (REML; $r = 0.47 \pm 0.11$) and heifers ($r = 0.58 \pm 0.09$). The distribution of scores differed between the bulls and heifers, the bulls being generally more fearful of human approach than the heifers. Sex differences in temperament measures have previously been found in cattle (Buchenauer, 1999). However, the difference seen here is likely to be due to environmental experience, as the heifers received much more human contact during rearing than the bulls.

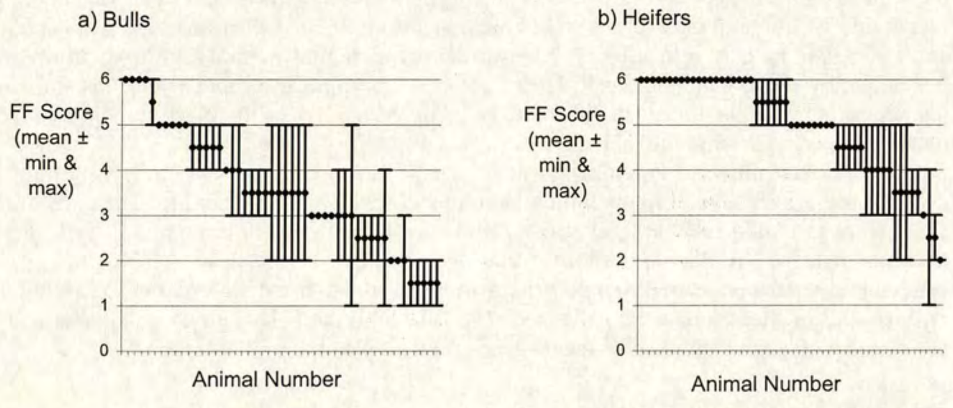


Figure 1a+b. Flight-from-Feeder scores obtained by a) the bull calves and b) the heifer calves

The score ranges from 1 (animal moves back when the observer is >2m away) to 6 (heifer doesn't move back when observer touches on the head).

MEASURING TEMPERAMENT TRAITS IN CATTLE FOR QTL IDENTIFICATION

N. Ball[1], M.J. Haskell[2], J.M. Deag[3] and J.L. Williams[1]

[1] Division of Genomics and Bioinformatics, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, Scotland, UK
[2] Animal Behaviour Sciences Dept., SAC, Bush Estate, Midlothian EH26 OPH, Scotland, UK
[3] Institute of Cell, Animal and Population Biology, The University of Edinburgh, Ashworth Laboratories, King's Buildings, West Mains Road, Edinburgh EH9 3JT, Scotland, UK

INTRODUCTION

Cattle show individual variation in their behavioural responses to handling and management systems on farms. These behavioural responses are presumed to reflect underlying temperament traits such as fear or aggression. Differences between breeds in behavioural traits demonstrate that variation in temperament has a genetic component, and heritabilities for such traits have generally been estimated as moderate to high (Burrows, 1997). Selection for favourable behavioural phenotypes would increase the ability of animals to cope with stressors encountered in modern agricultural systems, improving animal welfare and productivity, and human safety when handling stock. Hence information about the genetic loci that influence temperament may be of use in selective breeding programmes to select for animals with temperaments better suited to their environment.

Behavioural data obtained in temperament tests must be shown to reflect underlying traits by demonstrating intra-animal repeatability, inter-animal variability and validity. The aims of this study were to obtain reliable temperament measures in cattle using behavioural tests, and to use this data to localise quantitative trait loci (QTLs) involved in temperament. Two behavioural tests were carried out on groups of bull and heifer calves, and the variability and repeatability of the results were examined. The data from the behavioural phenotypes will be correlated with genetic markers to localise the QTLs involved.

MATERIALS AND METHODS

Behavioural responses to two important aspects of handling, human approach and separation from the group, were measured. A Flight-from-Feeder (FF) Test and a Social Separation (SS) Test were carried out on 49 bull calves and 56 heifer calves at approximately 11 months of age. The animals were second generation cross-bred Charolais x Holstein from a resource herd, and included F_2 calves and backcrosses to each parental breed. These animals were the first group of the 400 second-generation offspring in the herd. The bulls were reared with their dams at pasture for the first six months and then housed in groups. The heifers were reared on a dairy bucket system from 24 hours after birth and grouped at two weeks.

The FF Test. This test measured fearfulness of human approach, by measuring how quickly an animal moved away from a feeder when an observer approached. Each animal was given a score between 1 and 6 from a categorical scale, 1 being the score obtained when the animal moved back when the observer was >2m away, and 6 being the score obtained if the animal didn't move back when the observer touched it on the head.

The SS Test. This test measured how an animal responded to separation from its group when placed alone in its home pen for 5 minutes. Behaviours displayed during this time were continuously recorded, and total durations calculated. Principle Components Analysis (PCA) was used to group behaviours of a common motivational background.
Each test was carried out twice on each animal. Restricted Maximum Likelihood (REML) was used to calculate repeatability values.

RESULTS AND DISCUSSION

The FF Test demonstrated a wide range of scores between individuals (Figure 1a+b). High inter-animal repeatability of test scores was seen for both the bulls (REML; $r = 0.47 \pm 0.11$) and heifers ($r = 0.58 \pm 0.09$). The distribution of scores differed between the bulls and heifers, the bulls being generally more fearful of human approach than the heifers. Sex differences in temperament measures have previously been found in cattle (Buchenauer, 1999). However, the difference seen here is likely to be due to environmental experience, as the heifers received much more human contact during rearing than the bulls.

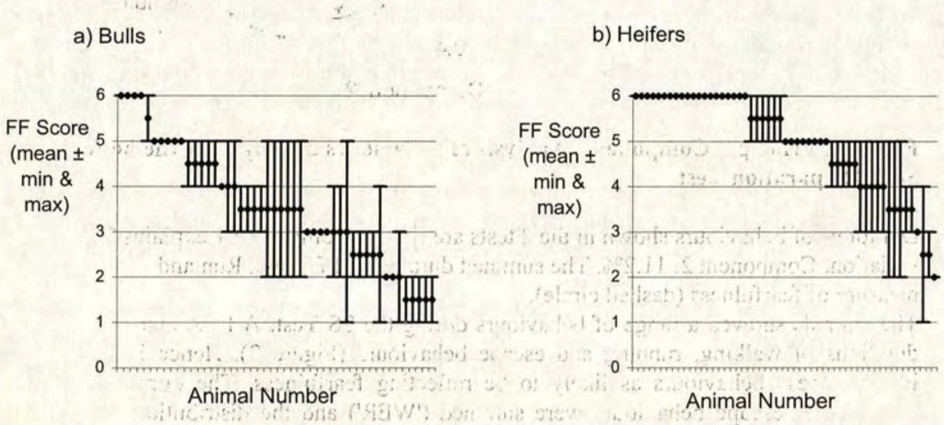


Figure 1a+b. Flight-from-Feeder scores obtained by a) the bull calves and b) the heifer calves

The score ranges from 1 (animal moves back when the observer is >2m away) to 6 (heifer doesn't move back when observer touches on the head).

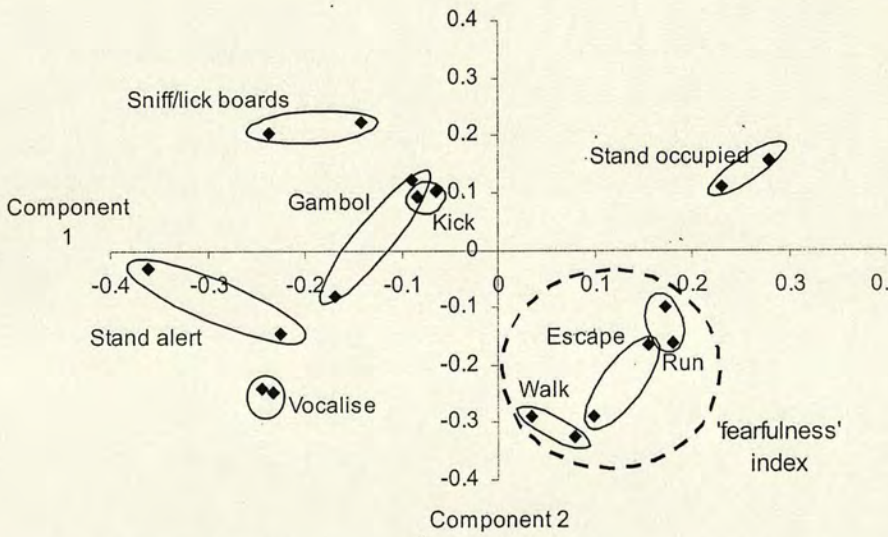


Figure 2. Principal Components Analysis of behaviours displayed by the heifers in the Social Separation Test

Durations of behaviours shown in the 2 tests are ringed. Component 1 explains 26.4% of the variation, Component 2, 11.2%. The summed durations of Escape, Run and Walk are used as a measure of fearfulness (dashed circle).

The animals showed a range of behaviours during the SS Test. A PCA analysis clustered the durations of walking, running and escape behaviours (Figure 2). Hence it was possible to identify these behaviours as likely to be reflecting fearfulness. The durations of walking, running and escape behaviours were summed ('WER') and the distributions of WER levels shown by the individual animals are shown in Figure 3a+b. A wide range in the level of these behaviours exhibited was found for both the bull and heifer calves. The heifers showed a high repeatability of $r = 0.63 \pm 0.11$, but the bulls had lower repeatability of $r = 0.40 \pm 0.13$. Again a difference in distribution was seen between the bull and heifer calves, with the bulls generally exhibiting lower levels of WER behaviours than the heifers. The heifers were kept in much closer proximity in indoor pens during rearing whereas the bulls were on grass for the first six months, and therefore differences between the sexes are again likely to be related to the rearing differences experienced by the animals.

CONCLUSION AND FUTURE PLANS

The data from the temperament tests demonstrated high inter-animal variability and intra-animal repeatability, from both the heifer and bull calves. Hence the data is suitable for use in

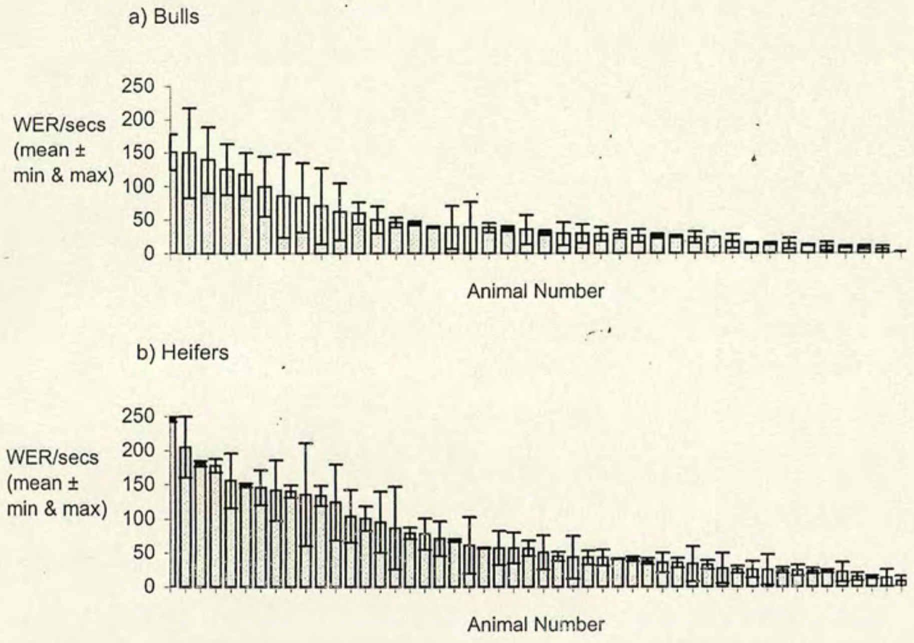


Figure 3. WER levels obtained in the SS Test by a) the bull calves and b) the heifer calves 'WER' is the combined total durations of the time spent walking, running and showing escape behaviours in 5 mins spent alone in the home pen.

associating behavioural phenotypes with genetic markers for the potential identification of QTLs. Sex differences in behavioural measurements were seen, but are likely to be due to differences in rearing methods. Regression methods for mapping QTLs are to be applied to the data and genotyping information from 186 microsatellite markers, which cover the genome at 20 cM intervals. Preliminary analysis of the transformed data from the F2 calves is underway, using QTL express (Seaton *et al.*; Haley *et al.*, 1994). Data on the analysis of the full number of animals will be presented, and should identify the major QTLs associated with temperament in cattle.

REFERENCES

Buchenauer, D. (1999) In « The Genetics of Cattle », p. 365-390, Editors R. Fries, and A. Ruvinsky, CAB International, Oxon, UK.

Burrow, H.M. (1997) *Anim. Breed. Abstr.* **65** : 477-495.

Haley, C.S., Knott, S.A. and Elsen, J.M. (1994). *Genetics* **136** : 1195-1207.

Seaton, G., Haley, C., Knott, S., Kearsey, M. and Visscher, P. QTL express http://qtl.cap.ed.ac.uk